



# **Análisis e interpretación de las medidas descriptivas con estudiantes de ciclo V mediante el trabajo por proyectos**

**Alejandro Albarracín Hernández**

Universidad Nacional de Colombia  
Facultad de Ciencias  
Maestría en Enseñanza de las Ciencias Exactas y Naturales  
Bogotá D.C., Colombia  
2019

# **Análisis e interpretación de las medidas descriptivas con estudiantes de ciclo V mediante el trabajo por proyectos**

**Alejandro Albarracín Hernández**

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título  
de:

**Magister en Enseña de las Ciencias Exactas y Naturales**

Director:

Dr. Luis Fernando Grajales Hernández

Universidad Nacional de Colombia

Facultad de Ciencias

Bogotá D.C., Colombia

2019

## *Dedicatoria*

*Este trabajo lo dedico a la persona, que me enseñó el verdadero sentido del amor.*

*La persona por la quien me levanto todos los días, quien es mi motor y mi motivación.*

*Gracias por existir.*

*Esto, es por ti*

*Te amo infinitamente*

*Juan José*

*Hijo mío.*

## **Agradecimientos**

Mis más sinceros agradecimientos al Dr. Luis Fernando Grajales Hernández por su continua colaboración, atención, dirección y asesoría en la ejecución de este trabajo de grado. De igual manera a todos los docentes de esta prestigiosa universidad que de manera desinteresada y con alto compromiso compartieron sus conocimientos.

Por otra parte, agradezco a las directivas del Colegio De La Salle Bogotá y especialmente a mis estudiantes de grado decimo, por su gran compromiso con las actividades desarrolladas en el presente trabajo.

Finalmente, agradezco a todas las personas que estuvieron a mi lado brindándome su apoyo, especialmente a ustedes Alejandra Rodríguez, Leonor Hernández, Carolina Albarracín, Felipe Albarracín y Luis Alejandro Albarracín por ser personas incondicionales quienes nunca se negaron a brindarme su apoyo y asesoría con mucha disposición. Para ellos infinitas gracias.

## Resumen

En el presente trabajo se elabora y se aplica una estrategia didáctica por medio del proyecto *Estadísticas de la prueba Saber año 2017*, que busca fortalecer y desarrollar mayores niveles de análisis, interpretación y comprensión de los conceptos de media aritmética, moda(s), mediana, varianza, desviación estándar, cuartiles y percentiles (medidas descriptivas), este fue desarrollado con 120 estudiantes de ciclo V (grado décimo) del Colegio de La Salle Bogotá. El propósito de este proyecto es el de dar respuesta a las dificultades que se han evidenciado en los estudiantes, en el análisis e interpretación de medidas descriptivas. Para su planeación se hizo énfasis en el trabajo por proyectos, por lo tanto, se diseñaron diferentes actividades en contexto a partir de datos de su interés, logrando así una mayor motivación, y participación. Los resultados de este trabajo fueron analizados de manera cualitativa y estadística haciendo un comparativo entre la prueba diagnóstica inicial y la prueba final, que fue aplicada posterior a la implementación del proyecto. Una vez analizado los resultados se evidenció un efecto positivo de la estrategia didáctica.

**Palabras clave:** medidas descriptivas (media, moda, mediana, varianza, desviación estándar, cuartiles y percentiles), trabajo por proyectos.

## Abstract

In the present work, a didactic strategy is developed and applied through the project Statistics of the “*pruebas saber*” from year 2017, which seeks to strengthen and develop higher levels of analysis, interpretation and comprehension of the concepts of arithmetic mean, mode, median, variance, standard deviation, quartiles and percentiles (descriptive measures). This work was developed with 120 students of cycle V (tenth grade) of La Salle School in Bogotá. The purpose of this project is to respond to the difficulties that have been evident in the students, in the analysis and interpretation of descriptive measures. For its planning, emphasis was placed on project work, therefore, different activities were designed in context based on data of interest, thus achieving greater motivation, and participation. The results of this work were analyzed qualitatively and statistically by comparing the initial diagnostic test and the final test, which was applied after the project was implemented. Once the results were analyzed, a positive effect of the didactic strategy was evident.

**Keywords:** descriptive measures (mean, mode, median, variance, standard deviation, quartiles and percentiles), project work.

# Contenido

	Pág.
<b>Resumen</b> .....	<b>IX</b>
<b>Abstract</b> .....	<b>X</b>
<b>Contenido</b> .....	<b>XI</b>
<b>Lista de figuras</b> .....	<b>XIII</b>
<b>Lista de tablas</b> .....	<b>XIV</b>
<b>Lista de Símbolos y abreviaturas</b> .....	<b>XVI</b>
<b>Introducción</b> .....	<b>1</b>
<b>1. Contextualización</b> .....	<b>3</b>
1.1. Ubicación geográfica y el contexto educativo .....	3
1.2. Planteamiento del problema .....	4
1.3. Pregunta de investigación .....	5
1.4. Objetivos .....	5
1.4.1. Objetivo general .....	5
1.4.2. Objetivos específicos .....	5
<b>2. Referentes teóricos</b> .....	<b>7</b>
2.1. Marco histórico y epistemológico .....	7
2.2. Marco disciplinar: Estadística descriptiva .....	10
2.2.1. Variables estadísticas .....	10
2.2.2. Tablas de frecuencia .....	12
2.2.2.1. Tablas de frecuencia para variables cualitativas .....	14
2.2.2.2. Tablas de frecuencia para una variable cuantitativa .....	15
2.2.3. Medidas de tendencia central (Moda, Mediana y Media) .....	16
2.2.3.1. Media Aritmética $\bar{x}$ .....	16
2.2.3.2. Mediana ( $Me$ ) .....	17
2.2.3.3. Moda ( $Mo$ ) .....	18
2.2.4. Medidas de posición .....	19
2.2.5. Medidas de dispersión .....	20
2.2.5.1. El rango .....	21
2.2.5.2. El rango intercuartílico .....	21
2.2.5.3. La varianza y la desviación estándar .....	22
2.2.6. Gráficos estadísticos .....	23
2.2.6.1. Gráficos para variables cualitativas .....	23
2.2.6.2. Gráficos para variables cuantitativas .....	24
2.3. Marco metodológico: Investigación – Acción .....	26
2.4. Marco didáctico: Trabajo por proyectos .....	28
<b>3. Estrategia didáctica</b> .....	<b>30</b>
3.1. Prueba diagnóstica .....	31
3.1.1. Análisis a priori de la prueba diagnóstica .....	31

3.1.2. Descripción general de las sesiones previas al desarrollo del proyecto.....	37
3.2. El proyecto: Las estadísticas de la prueba saber año 2017. ....	39
3.2.1. Objetivos del proyecto .....	40
3.2.2. Los datos.....	40
3.2.3. Desarrollo de la actividad .....	41
Análisis a priori de las preguntas del proyecto.....	41
<b>4. Estrategia didáctica: Resultados y análisis .....</b>	<b>52</b>
4.1. Prueba diagnóstica: Análisis de resultados .....	52
4.1.1. Análisis cualitativo, prueba diagnóstica, previo a la implementación del proyecto.....	53
4.1.2. Análisis cualitativo, prueba final, después de implementar el proyecto.....	56
4.1.3. Análisis estadístico de la prueba diagnóstica, antes y después de aplicar el proyecto Estadísticas de la prueba Saber 2017. ....	58
4.1.3.1. Conclusiones por pregunta. ....	59
4.2. Proyecto Estadísticas de las pruebas Saber año 2017. Análisis de resultados .	65
<b>5. Conclusiones y recomendaciones .....</b>	<b>71</b>
5.1. Conclusiones sobre los objetivos .....	71
5.2. Recomendaciones .....	74
<b>6. BIBLIOGRAFIA.....</b>	<b>87</b>



## Lista de figuras

	Pág.
Figura 2-1 Ejemplos de diagrama de cajas y bigotes. (Boxplot) .....	25
Figura 3-1 Boxplot de la variable puntaje de matemáticas.....	49
Figura 3-2 Histograma de la variable edad V3.....	50
Figura 3-3 Polígono de frecuencias de la variable edad V3.....	51
Figura 4-1 Resultados generales en las preguntas una vez aplicada la “prueba final”...	64
Figura 4-2. Ejemplo de respuesta ítem 1 .....	66
Figura 4-3. Ejemplo 2 respuesta de ítem 1.....	66
Figura 4-4. Ejemplo de respuesta ítem 2.....	67
Figura 4-5. Ejemplo de respuesta ítem 3.....	67
Figura 4-6. Ejemplo de respuesta ítem 4.....	68
Figura 4-7 Ejemplo de boxplot ítem 5.....	68
Figura 4-8 Ejemplo 2 boxplot ítem 5.....	69
Figura 4-9. Ejemplo 3 boxplot ítem 5.....	69
Figura 4-10. Ejemplo 1 Histograma y polígono de frecuencias ítem 6.....	70

## Lista de tablas

	Pág.
<b>Tabla 2.1</b> Formato de una tabla de distribución de frecuencias .....	13
<b>Tabla 2.2</b> Encuesta de opinión acerca de una medida económica.....	14
<b>Tabla 2.3</b> Distribución de frecuencias de la tabla 2.2.....	14
<b>Tabla 3.1</b> Esquema propuesto por el ICFES para el análisis conceptual y procedimental .....	31
<b>Tabla 3.2</b> Análisis conceptual y procedimental pregunta 1.....	32
<b>Tabla 3.3</b> Análisis conceptual y procedimental pregunta 2.....	33
<b>Tabla 3.4</b> Análisis conceptual y procedimental pregunta 3.....	33
<b>Tabla 3.5</b> Análisis conceptual y procedimental pregunta 4.....	33
<b>Tabla 3.6</b> Análisis conceptual y procedimental pregunta 5 .....	34
<b>Tabla 3.7</b> Análisis conceptual y procedimental pregunta 6 .....	34
<b>Tabla 3.8</b> Análisis conceptual y procedimental pregunta 7 .....	35
<b>Tabla 3.9</b> Análisis conceptual y procedimental pregunta 8 .....	35
<b>Tabla 3.10</b> Análisis conceptual y procedimental pregunta 9.....	36
<b>Tabla 3.11</b> Análisis conceptual y procedimental pregunta 10 (a) .....	36
<b>Tabla 3.12</b> Análisis conceptual y procedimental pregunta 10 (b) .....	37
Tabla 3.13 Datos de la encuesta Habitantes de calle Bogotá D.C. año 2017.....	38
Tabla 3.14 Datos de Estadísticas vitales Colombia año 2017. ....	38
<b>Tabla 3.15</b> Datos de los resultados de la prueba saber año 2017 .....	40
Tabla 3.16 Ítem 1 actividad del proyecto.....	41
Tabla 3.17 Tipos de variables del grupo A. ....	42
Tabla 3.18 Ítem 2 del proyecto.....	42
Tabla 3.19 Tabla de distribución de frecuencias de la variable V3 .....	42
Tabla 3.20 Ítem 3 del proyecto.....	44
Tabla 3.21 Ítem 4 del proyecto .....	45
Tabla 3.22 Tabla de distribución de frecuencias de la variable V8. ....	45
Tabla 3.23 Tabla de distribución de frecuencias de la variable V10 .....	46
Tabla 3.24 Ítem 5 del proyecto.....	48
Tabla 3.25 Ítem 6 del proyecto.....	49

---

Tabla 3.26 Distribución de frecuencias de datos agrupados para V3 .....	50
Tabla 4.1 Resultados de la prueba diagnóstica grupo 1001 .....	54
Tabla 4.2 Resultados de la prueba diagnóstica grupo 1002 .....	54
Tabla 4.3 Resultados de la prueba diagnóstica grupo 1003 .....	54
Tabla 4.4 Resultados de la prueba diagnóstica grupo 1004 .....	54
<b>Tabla 4.5</b> Análisis cualitativo de las dificultades en el desarrollo de la prueba diagnóstica. .....	55
Tabla 4.6 Resultados de la prueba final grupo 1001.....	56
<b>Tabla 4.7</b> Resultados de la prueba final grupo 1002. ....	56
<b>Tabla 4.8</b> Resultados de la prueba final grupo 1003. ....	57
<b>Tabla 4.9</b> Resultados de la prueba final grupo 1004. ....	57
Tabla 4.10. Presentación de los datos.....	58
Tabla 4.11. Datos de la prueba. Pregunta 1. ....	59
Tabla 4.12. Datos de la prueba. Pregunta 2. ....	59
Tabla 4.13. Datos de la prueba. Pregunta 3. ....	60
Tabla 4.14. Datos de la prueba. Pregunta 4. ....	60
Tabla 4.15. Datos de la prueba. Pregunta 5. ....	60
Tabla 4.16. Datos de la prueba. Pregunta 6. ....	61
Tabla 4.17. Datos de la prueba 1001 pregunta 7.....	61
Tabla 4.18. Datos de la prueba. Pregunta 8. ....	62
Tabla 4.19. Datos de la prueba. Pregunta 9. ....	62
Tabla 4.20. Datos de la prueba. Pregunta 10a. ....	62
Tabla 4.21. Datos de la prueba. Pregunta 10b. ....	63

# Lista de Símbolos y abreviaturas

## Símbolos con letras latinas

Símbolo	Término
$f$	Frecuencia absoluta
$F$	Frecuencia absoluta acumulada
$h$	Frecuencia relativa
$H$	Frecuencia relativa acumulada
$R_x$	Rango
$x_{min}$	Valor mínimo de un conjunto de datos
$x_{max}$	Valor máximo de un conjunto de datos
$A$	Amplitud
$\bar{x}$	Media aritmética
$Me$	Mediana
$Mo$	Moda
$P_k$	Percentil
$S^2$	Varianza

## Abreviaturas

Abreviatura	Término
ICFES	Instituto Colombiano para el Fomento de la Educación Superior
<i>MEN</i>	Ministerio de Educación Nacional

# Introducción

Si bien la estadística en Colombia se enseña en todos los niveles de la primaria y la secundaria, de acuerdo a lo establecido en los lineamientos curriculares y estándares básicos de competencias en matemáticas propuesto por el MEN. (Ministerio de Educación Nacional) Es notoria, desde la práctica docente la dificultad que presentan algunos estudiantes al momento de aplicar los conceptos y procedimientos estadísticos.

(Batenero , Díaz, Contreras, & Roa, 2013) Manifiestan, que es de gran importancia la enseñanza de la estadística; esto se debe a la necesidad de varias instituciones como la UNESCO por fomentar una *cultura estadística* que les permita a los ciudadanos participar en la sociedad de la información. Esta cultura se hace necesaria en la lectura de algunos artículos de prensa, en la interpretación de información propuesta en diferentes medios (internet, vallas publicitarias, noticieros...), en la participación de encuestas e incluso en la interpretación de algunos diagnósticos médicos.

La siguiente propuesta se enmarca en el proceso de enseñanza - aprendizaje de la estadística y pretende proponer una estrategia didáctica que fortalezca y desarrolle mayores niveles de análisis, interpretación y comprensión de los conceptos de media aritmética, moda(s), mediana, varianza, desviación estándar, cuartiles y percentiles (**medidas descriptivas**). Para el desarrollo de este propósito se utilizará como metodología, el trabajo con proyectos, ya que como menciona Watson citado en (Batanero & Díaz , Estadística con proyectos, 2011), es importante que los alumnos se enfrenten a problemas estadísticos en donde el contexto juegue un papel importante y sea de interés para ellos ya que esto facilitará la comprensión de los distintos contenidos.

En un apartado del documento de Los lineamientos curriculares en Matemáticas se muestra la subdivisión de los pensamientos en cinco tipos: el numérico, el espacial, el métrico o de medida, el variacional y el aleatorio o probabilístico. Este último es donde estará apuntando nuestra investigación. Específicamente a lo propuesto en los

Estándares Básicos de Competencia en Matemáticas en donde se afirma que *“al terminar undécimo grado... Usa comprensivamente algunas medidas de centralización, localización, dispersión y correlación (percentiles, cuartiles, centralidad, distancia, rango, varianza, covarianza y normalidad)”*. (MEN, 2006)

El desarrollo de este proyecto surge a partir de un interés propio por trabajar con líneas de investigación en la enseñanza de la estadística y por generar en los estudiantes habilidades estadísticas que contribuyan a su desarrollo personal y laboral. El estudio se desarrollará en el Colegio De La Salle Bogotá, colegio que cuenta en la actualidad con más de 800 estudiantes en jornada única y se realizará con 120 estudiantes de grado décimo.

# **1. Contextualización**

Durante el desarrollo de este capítulo se abordarán aspectos como: La ubicación geográfica, el contexto institucional y escolar donde se desenvuelve la propuesta, el planteamiento del problema, la pregunta de investigación y los objetivos propuestos.

## **1.1. Ubicación geográfica y el contexto educativo**

El Colegio De La Salle, donde se desarrolló la propuesta está ubicado en la localidad de Usaquén al nororiente de Bogotá D.C. Es una institución Educativa mixta de carácter privado, que ofrece dentro de su proyecto educativo institucional (PEI) un alto nivel en formación académica y un énfasis en la formación de valores cristianos y católicos bajo la filosofía de San Juan Bautista De La Salle.

Fue fundado por la comunidad de los Hermanos de La Salle, quienes desde hace tres siglos se dedican a la formación y educación de niños y jóvenes en el Mundo y desde 1890 ejercen la labor educativa en Colombia, actualmente la comunidad dirige 44 obras educativas en todo el territorio nacional, de los cuales 7 colegios quedan en Bogotá D.C., dentro de ellos el Colegio De La Salle; que fue fundado hace 43 años y ha procurado que los principios en la formación lasallista mixta sean de acuerdo al horizonte educativo pastoral, una apuesta centrada en la enseñanza humanista y espiritual. (La Salle Distrito Bogotá , 2019)

La población del colegio se caracteriza por tener estratos socioeconómicos entre 4,5 y 6, donde la principal problemática identificada es la falta de acompañamiento por parte de los padres de familia, debido a sus trabajos ya que en ocasiones trabajan padre y madre o también por la administración de sus diferentes empresas.

La propuesta se realizó con 120 estudiantes que están divididos en 4 cursos, los estudiantes pertenecen al ciclo V, más específicamente al grado décimo, cuyo rango de edad esta entre los 15 y 18 años.

## 1.2. Planteamiento del problema

Para el desarrollo de esta investigación se optó por considerar temas en estadística específicamente los temas de medidas descriptivas, en particular los conceptos de media, moda(s), mediana, varianza, desviación estándar, cuartiles, deciles y percentiles, debido a la importancia que tienen estos en el análisis e interpretación de datos. Pero también por la importancia que tiene la estadística en cualquier área del saber, autores como Iddo Gal, plantean el término “Statistical Literacy” o alfabetización estadística para destacar el hecho que hoy en día la estadística se considera parte fundamental para el ciudadano educado, (Batenero , Díaz, Contreras, & Roa, 2013).

Para el MEN es de suma importancia que los estudiantes obtengan una buena formación en estadística, esto se hace evidente en los documentos base del currículo escolar colombiano como los son los Lineamientos Curriculares de Matemáticas, los Estándares Básicos de Competencias en Matemáticas y los Derechos Básicos de aprendizaje (DBA) en donde se establecen cómo y cuáles son los contenidos que se desarrollarán en cada uno de los ciclos establecidos. Por ejemplo, en los DBA (MEN, 2016) de matemáticas para décimo grado establece el DBA número 9: *“Comprende y explica el carácter relativo de las medidas de tendencias central y de dispersión, junto con algunas de sus propiedades, y la necesidad de complementar una medida con otra para obtener mejores lecturas de los datos.”*

En el desarrollo de las clases se ha notado las dificultades que presentan los estudiantes cuando se enfrentan a situaciones que requieren de conocimiento estadístico. Por ejemplo, situaciones en donde es necesario describir, juntar, analizar o relacionar las medidas descriptivas, para realizar una mejor lectura de los datos o también cuando deben realizar interpretaciones o analizar información resultante de gráficas como por ejemplo histogramas o boxplot, provenientes de diferentes medios de comunicación. Este tipo de dudas y otras que se han manifestado por parte de los estudiantes muestran que hay desconocimiento de los conceptos y de sus propiedades; además, es importante tener en cuenta que mucha información que se presenta en los medios requiere del conocimiento de estos conceptos, por eso la importancia que los estudiantes tengan claridad sobre estos. De acuerdo con lo expuesto anteriormente aparece la iniciativa de contribuir en el desarrollo de la cultura estadística que brinde las habilidades necesarias, que les permita



a los ciudadanos (estudiantes) participar, interpretar y evaluar críticamente algún tipo de información presentada.

### 1.3. Pregunta de investigación

De acuerdo a la problemática planteada anteriormente y trabajando en el desarrollo de los procesos planteadas en los Lineamientos Curriculares en Matemáticas (MEN, 1998, págs. 52-82) surge la siguiente pregunta:

¿Qué estrategia didáctica es la adecuada para trabajar con los estudiantes de grado décimo las temáticas de medidas descriptivas?

### 1.4. Objetivos

#### 1.4.1. Objetivo general

Desarrollar e implementar una estrategia didáctica para trabajar con los estudiantes de grado décimo las medidas descriptivas mediante la metodología del trabajo por proyectos.

#### 1.4.2. Objetivos específicos

- ✓ Identificar los conocimientos previos de los estudiantes con respecto a los conceptos de media aritmética, modas(s), mediana, varianza, desviación estándar, cuartiles y percentiles (**Medidas descriptivas**).
- ✓ Proponer ejercicios que conduzcan a la utilización de hojas de cálculo Excel en el análisis de datos, estadística descriptiva.
- ✓ Seleccionar la información, los contextos y las situaciones idóneas para diseñar las actividades que den estructura a la estrategia didáctica mediante el trabajo por proyectos
- ✓ Implementar y evaluar la estrategia didáctica.

## 2. Referentes teóricos

Durante el desarrollo de este capítulo, se aborda el marco histórico y epistemológico donde se hace mención de aspectos relevantes que marcaron el desarrollo de la estadística. Posteriormente en el marco disciplinar se realiza una síntesis de conceptos de estadística descriptiva entorno a los conceptos de medidas descriptivas, puntualmente los conceptos de medidas de tendencia central, medidas de posición y de dispersión; algunos gráficos estadísticos como lo son, el boxplot, histograma y polígono de frecuencias.

En el marco didáctico se desarrolla una breve descripción del trabajo con proyectos en donde se recopilan los lineamientos básicos para la implementación de la metodología de aprendizaje y por último se realiza una descripción de las diferentes fases que se abordaron de la metodología de investigación acción.

### 2.1. Marco histórico y epistemológico

Realizando una búsqueda del concepto de estadística, se encontraron múltiples definiciones y entre ellas hemos elegido la siguiente, que según Carmen Batanero reúne muy bien la concepción al respecto (Batanero, Didáctica de la Estadística , 2001)

*"La estadística estudia el comportamiento de los fenómenos llamados de colectivo. Está caracterizada por una información acerca de un colectivo o universo, lo que constituye su objeto material; un modo propio de razonamiento, el método estadístico, lo que constituye su objeto formal y unas previsiones de cara al futuro, lo que implica un ambiente de incertidumbre, que constituyen su objeto o causa final."* Cabriá citado en (Batanero, Didáctica de la Estadística , 2001) .

La aparición de la estadística se debe a la necesidad de los diferentes Gobiernos, por conocer hechos políticos, económicos o sociales para establecer sus riquezas, el número de habitantes, su producción, sus aspectos demográficos y fiscales; estas

técnicas iniciales se conocen como estadística descriptiva, caracterizada por procesos de conteo, recolección de datos, descripción y análisis de la información. (Hidalgo, 2016)

Las referencias en cuanto a la recolección y conteo de datos se remontan a tiempos muy antiguos, podría decirse que la historia comienza en la isla de Cerdeña donde se identifican algunos monumentos prehistóricos donde se encuentran ciertas cisuras que permiten suponer un conteo que se hacía sobre el ganado cazado. La existencia de la estadística se debe a la necesidad de las sociedades de conocer hechos económicos, políticos, etc. En relación con el dominio de territorios, riquezas y la producción agrícola, Carrasco (citado en Hurtado, 2013)

La estadística en sus orígenes era rudimentaria y se relacionaba con registros y censos de campo, enfocados a lo económico, lo social y lo educacional. Se tiene conocimiento que los babilónicos, utilizaban tablillas hechas en arcilla para recopilar datos acerca de la población, la producción agrícola y cambios hechos por trueque. En la antigua china Confucio en su texto Shu-King, narra la historia del rey Yao en el año 2238 a.C. quien ordenó que se hiciera una recopilación de datos acerca de la producción agrícola, industrial y comercial, como también la realización de censos para calcular los impuestos. (Hurtado et al., 2013)

En la biblia también se reconocen registros estadísticos como, por ejemplo, el censo ordenado por el rey David a Joab, general del ejército para conocer el número de habitantes de la población o el también realizado por Moisés después de la salida de Egipto.

#### ***“Censo de Israel en Sinaí***

*1Habló Jehová a Moisés en el desierto de Sinaí, en el tabernáculo de reunión, en el día primero del mes segundo, en el segundo año de su salida de la tierra de Egipto, diciendo: 2Tomad el censo de toda la congregación de los hijos de Israel por sus familias, por las casas de sus padres, con la cuenta de los nombres, todos los varones por sus cabezas. 3De veinte años arriba, todos los que pueden salir a la guerra en Israel, los contaréis tú y Aarón por sus ejércitos. 4Y estará con*

*vosotros un varón de cada tribu, cada uno jefe de la casa de sus padres.”* (Números 1: 1-4)

Durante el tiempo del Imperio Romano se favoreció el desarrollo y empleo de recursos estadísticos, ya que cada cinco años se realizaban censos de población y se registraban los nacimientos, defunciones y matrimonios, como también se llevaba recuento del ganado, de las riquezas que se poseían en cada uno de los territorios conquistados. Los siguientes mil años después de la caída del Imperio Romano se tiene poco conocimiento de otro tipo de operaciones estadísticas, aunque existe referencia de las tierras poseídas por parte de la iglesia, que fueron compiladas por Carlomagno y Pipino el Breve. (Estrella A. , 2011)

La evolución histórica de los conceptos **de medidas tendencia central** (Media, Mediana y moda) se da inicialmente por la noción del “hombre promedio” (l’homme moyen) introducida por Quetelet quien fuese otro importante fundador de la estadística. (Estrella S. , 2008) Al respecto batanero en el año 2000: *“El surgimiento progresivo del concepto que hoy conocemos como “media aritmética”, comenzó como útil implícito en la solución de problemas prácticos, y más tarde como objeto de estudio en sí mismo. Los problemas prácticos y más tarde teóricos, han llevado a la definición del concepto de media, a la identificación de sus propiedades, más tarde a la definición de otras medidas de posición central, como la mediana o moda, que son preferibles a la media en algunas situaciones concretas. Además, ha sido necesario „probar“ o „demostrar“ la validez de estas soluciones y propiedades, para aceptarlas como parte del conocimiento matemático.”*

Por otra parte, con respecto a la evolución histórica de las **medidas de dispersión** nos apoyaremos en el artículo de Pearson, Kendall y Plackett, (citado en Ruiz, 2017) en donde se menciona que los estudios realizados por astrónomos babilonios, entre los años 500 y 300 a. C., en el que predecían el movimiento del sol y de otros astros, se podía establecer como surgía la media, pero implícito en este problema surge el concepto de dispersión, ya que para desarrollar este tipo de observaciones era necesario una correlación entre predicción y observación, aunque no se tiene información de como hacían para resolverlo. En el mismo artículo presentan el caso de Hiparco quien investigó el paso del sol en el solsticio por un mismo punto y se cuestionó sobre si el año tropical es constante o no; durante su investigación el cuantifico su error que correspondía a un cuarto de día, es decir

utilizó una medida de dispersión para caracterizar el error. Sin embargo, hasta el año 1600 Galileo realizó la primera descripción matemática del error en donde se describe la distribución del error como simétrica y como circundante a un valor verdadero (ya se intuía el concepto de media), dicha contribución fue catalogada como la principal contribución a la teoría de la probabilidad por parte de Galileo.

Una característica que se da en la histórica de las matemáticas, es que esta son un conjunto de conocimientos que están en continua evolución y que durante esta evolución desempeña un papel de primer orden la necesidad de resolver determinados problemas prácticos o propios de las mismas matemáticas y su interrelación con otros conocimientos (otras ciencias). (Didáctica de las Matemáticas para Maestros, 2004)

## 2.2. Marco disciplinar: Estadística descriptiva.

Para el desarrollo del marco disciplinar se utilizan los textos *Estadística Descriptiva* (Monroy Saldívar, 2008), *Estadística Descriptiva y Probabilidades* (Sosa Martínez, Ospina Forero, & Berdugo Camacho, 2013) y trabajos realizados en la MEC como el de (Chaparro, 2018) y (Rojas, 2017) y aportes de reuniones con el profesor Luis Fernando Grajales Hernández, director del presente trabajo.

La estadística descriptiva según, Sosa Martínez, Ospina Forero, & Berdugo Camacho (2013) “describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y representan la información contenida en ellos. “A continuación haremos una breve descripción de algunos términos importantes en el desarrollo de este trabajo.

### 2.2.1. Variables estadísticas

Las variables tienen un papel fundamental en el desarrollo de una investigación estadística, ya que lo primero que se determina es el asunto a investigar para poder determinar las variables que intervienen.

“Una **variable** es un característica observable o medible de un individuo, que se describe acorde a una escala de medición bien definida. Cada rasgo, aspecto o

*característica de una población constituye una variable”* (Sosa Martínez et al., 2013, p. 7)

Las variables se clasifican en cualitativas o cuantitativas y dependiendo de su clasificación se determina su escala categórica o numérica.

Las **variables cualitativas** se expresan como categorías o atributos y su escala es de tipo nominal u ordinal.

**Escala nominal**, se clasifican las unidades de estudio en categorías teniendo en cuenta la cantidad de características. A cada categoría le corresponde un nombre que puede ser alfabético o alfanumérico, cabe aclarar que en esta categoría los números no se pueden manipular aritméticamente puesto que cumplen una función únicamente de asignación. Adicionalmente, las variables de tipo nominal se pueden clasificar en dicotómicas cuando incluyen únicamente dos categorías, o politómicas cuando contiene tres o más categorías.

**Escala ordinal**, sus valores representan categorías con alguna clasificación jerárquica específica. Una vez clasificados, es posible determinar qué objeto tiene más o menos de una característica en comparación con otro, pero no es posible establecer en cuánto difieren.

Las **variables cuantitativas** expresan una cantidad numérica, se clasifican en discretas y continuas.

**Variables discretas** son aquellas que en sus valores están separadas entre sí por una determinada cantidad, una característica de las variables discretas es la presencia de “vacíos” o “interrupciones”. (Monroy Saldívar, 2008)

**Variables continuas**, son aquellas que pueden tomar cualquier valor dentro de un intervalo de valores.

Las variables cuantitativas utilizan escalas numéricas de intervalo o de razón. En una **escala de intervalo** los elementos se ordenan por la magnitud del atributo que está representado y utiliza intervalos de igual magnitud; no poseen un cero absoluto, dado que se establece por convención de forma arbitraria por los expertos del área de estudio; el

cero no implica la ausencia del atributo. En la **escala de razón** el punto cero no es arbitrario y corresponde a una ausencia de la característica observada. (Sosa Martínez, Ospina Forero, & Berdugo Camacho, 2013)

### 2.2.2. Tablas de frecuencia

Una de las primeras etapas en el análisis descriptivo es la recopilación de datos en tablas, en donde se resume la información de las variables de manera compacta y precisa. Al respecto Sosa Martínez, Ospina Forero, & Berdugo Camacho mencionan que las **tablas de frecuencia**, se elaboran teniendo en cuenta un conjunto de  $n$  elementos asociado con una variable cuyos valores se van agrupando en categorías que denominan  $k$  o también con intervalos que están denotados de la siguiente manera  $C_1, C_2, C_3, \dots, C_k$ . Para cada una de las  $k$  clases se definen las siguientes magnitudes.

La **frecuencia absoluta**, de la clase  $C_i$ , denotada  $f_i$  corresponde a la cantidad de observaciones que pertenecen  $i$ -ésima categoría o intervalo para  $i = 1, \dots, k$ .

La **frecuencia relativa** de la clase  $C_i$ , denotada  $h_i$  es la proporción del  $f_i$  respecto al total de observaciones, es decir.

$$h_i = \frac{f_i}{n} \quad (2-1)$$

Para  $i = 1, \dots, k$ .

La **frecuencia absoluta acumulada** de la clase  $C_i$ , denotada  $F_i$ , corresponde a la cantidad de individuos cuya modalidad es inferior o igual a la  $i$ -ésima categoría.

$$F_i = \sum_{j=1}^i f_j \quad (2-2)$$

Para  $i = 1, \dots, k$ .

La **frecuencia relativa acumulada** de la clase  $C_i$ , denotada  $H_i$ , corresponde a la proporción de individuos respecto al total de observaciones cuya modalidad es inferior o igual la  $i$ -ésima categoría.

$$H_i = \frac{F_i}{n} \quad (2-3)$$

Para  $i = 1, \dots, k$ .

Se llama **distribución de frecuencias** a la matriz que sirve para representar de forma organizada una distribución de los datos, dicha distribución contiene todas las categorías junto con sus correspondientes frecuencias. La **Tabla 2.1** Formato de una tabla de distribución de frecuencias muestra la correspondiente distribución.

Es importante tener en cuenta, que en una tabla de distribución de una categoría con  $k$  clases se pueden verificar las siguientes propiedades:

$$\text{i.} \quad \sum_{i=1}^k h_i = 1 \quad (2-4)$$

$$\text{ii.} \quad F_k = n \quad (2-5)$$

$$\text{iii.} \quad H_i = \sum_{i=1}^k h_k \quad (2-6)$$

$$\text{iv.} \quad H_k = 1 \quad (2-7)$$

**Tabla 2.1** Formato de una tabla de distribución de frecuencias

$c_i$	$f_i$	$F_i$	$h_i$	$H_i$
$c_1$	$f_1$	$F_1$	$h_1$	$H_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$c_k$	$f_k$	$n$	$h_k$	$1$
<b>TOTAL</b>	$n$			



### 2.2.2.1. Tablas de frecuencia para variables cualitativas

En la construcción de tablas estadísticas se debe tener en cuenta la escala de medición, para determinar si es acertado o no emplear las frecuencias acumuladas. En el caso de las variables de tipo nominal no tiene sentido las frecuencias acumuladas. A continuación, se enuncia un ejemplo tomado de Sosa Martínez, Ospina Forero, & Berdugo Camacho (2013)

Considere el conjunto de datos de la **Tabla 2.2** Encuesta de opinión acerca de una medida económica asociados a una encuesta de opinión acerca de medida económica. Elabore la tabla de frecuencias correspondiente.

**Tabla 2.2** Encuesta de opinión acerca de una medida económica

1	0	0	1	1	3	1
3	1	3	0	0	0	3
0	0	0	1	1	0	1
0	0	0	0	0	1	1

La **Tabla 2.2** Encuesta de opinión acerca de una medida económica. Datos asociados con una encuesta de opinión acerca de una medida económica. 1=A favor; 0=En contra; 3= No sabe / No responde (NS/NR), De la anterior información podemos concluir que el número de clases  $k = 3$ . La **Tabla 2.3** Distribución de frecuencias de la tabla 2.2

**Tabla 2.3** Distribución de frecuencias de la tabla 2.2

$c_i$	$f_i$	$h_i$
A favor	10	35.7 %
En contra	14	50.0 %
NS/NR	4	14.3 %
TOTAL	28	100 %

Como la escala de medición de la variable es de “opinión”, es decir, nominal; entonces no tiene sentido incluir las frecuencias acumuladas.

### 2.2.2.2. Tablas de frecuencia para una variable cuantitativa.

En el estudio de las variables cuantitativas se puede resumir la información como se presentó en la **Tabla 2.1** Formato de una tabla de distribución de frecuencias. Sin embargo cuando la variable es cuantitativa de escala de razón, no se dispone de un conjunto  $k$  clases para elaborar la tabla de distribución de frecuencias. El siguiente es un procedimiento para crear las categorías para este tipo de variable utilizado en (Sosa Martínez, et al., 2013).

Hay varias maneras de determinar la cantidad  $k$  de intervalos (clases) de agrupación un caso es utilizar.

$$k = \lceil \sqrt{n} \rceil \quad (2-8)$$

Donde  $n$  es el tamaño de la muestra y se en donde se define la función techo como el menor entero mayor igual que el número dado, su notación es  $\lceil \cdot \rceil$ . Por ejemplo,

$$\lceil 2.01 \rceil = 3$$

$$\lceil 12.5 \rceil = 13$$

Otra opción es utilizar la **regla de Sturges**. Que está dada por la ecuación.

$$k = \lceil 1 + 3.32 \log_{10}(n) \rceil \quad (2-9)$$

Donde  $n$  es el tamaño de la muestra.

**El rango** para una variable  $X$ , denotado  $R_X$ , es un valor resultante entre la diferencia del valor máximo ( $x_{max}$ ) y el valor mínimo ( $x_{min}$ )

$$R_X = x_{max} - x_{min} \quad (2-10)$$

Para determinar la **amplitud**  $A$  de categorías, se utiliza la siguiente expresión, donde  $k$  es la cantidad de intervalos.

$$A = \frac{R_X}{k} \quad (2-11)$$

Las categorías se establecen de la siguiente manera:

$$C_1 = \{x: l_0 \leq x < l_1\} \text{ donde } l_0 = x_{min} \text{ y } l_1 = l_0 + A$$

$$C_2 = \{x: l_1 \leq x < l_2\} \text{ donde } l_2 = l_1 + A$$

$$\vdots$$

$$C_k = \{x: l_{k-1} \leq x < l_k\} \text{ donde } l_k = l_{k-1} + A = x_{max}$$

### 2.2.3. Medidas de tendencia central (Moda, Mediana y Media)

Una medida de tendencia central es un único valor que indica el centro de una serie de números (Monroy Saldívar, 2008). Al respecto Sosa Martínez, Ospina Forero, & Berdugo Camacho (2013) establecen que cuando se realiza un análisis descriptivo de las variables de estudio se determinan características relevantes de la información, ahora se indaga por las propiedades de la distribución de los datos. De este modo se puede observar una tendencia de los datos a agruparse alrededor de los valores más frecuentes. De aquí surge la necesidad de las medidas de tendencia central. Tales medidas son importantes en el análisis de la información, sin embargo, su interpretación no debe desligarse de las medidas de dispersión ya que la calidad de las medidas de tendencia central está asociada intrínsecamente con el grado de concentración de la información.

Para el cálculo de las medidas de tendencia central partiremos de un conjunto de datos que denotaremos  $x_1, x_2, x_3, x_4, \dots, x_n$  siendo  $n$  la cantidad total de datos del conjunto.

#### 2.2.3.1. Media Aritmética ( $\bar{x}$ )

##### Datos sin agrupar

Si los datos de una muestra están ordenados en una tabla de frecuencias la media aritmética o promedio, se determina como la suma de las observaciones de los datos del conjunto, ponderada con la frecuencia de los mismos; dividida entre la cantidad total de datos del conjunto.

$$\bar{x} = x_1 h_1 + \dots + x_k h_k = \frac{1}{n} (x_1 f_1 + \dots + x_k f_k)$$

$$= \frac{1}{n} \sum_{i=1}^k x_i f_i \quad (2-12)$$

Si los datos no están ordenados en una tabla, entonces.

$$\bar{x} = \frac{x_1 + \dots + x_n}{n} \quad (2-13)$$

### Datos agrupados

Cuando los datos de una muestra están agrupados en una distribución de frecuencias la media aritmética se calcula con la siguiente formula.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k f_i X'_i \quad (2-14)$$

Donde  $X'_i$  corresponde a la marca de clase de la  $i$ -ésima categoría.

### 2.2.3.2. Mediana ( $Me$ )

Es una medida de posición y se define como el número real  $Me$  tal que al menos el 50% de los datos es menor o igual que  $Me$  y al menos el 50% de los datos es mayor o igual que  $Me$ . Se debe calcular una vez los datos estén ordenados. (Chaparro, 2018)

### Datos sin agrupar

Dados los datos  $x_1 \leq x_2 \leq x_3 \leq x_4, \dots, \leq x_n$  de una variable  $x_i$ ; donde  $x_i$  es el dato que ocupa la  $i$ -ésima posición. La Mediana se calcula como sigue:

Si  $n$  corresponde a una cantidad impar de datos, entonces.

(2-15)

$$Me = x_{\left(\frac{n+1}{2}\right)}$$

Es decir el valor de la posición  $\frac{n+1}{2}$

Si  $n$  corresponde a una cantidad par de datos, entonces.

(2-16)

$$Me = \frac{1}{2} \left( x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)} \right)$$

Es decir, la semisuma de los valores de los dos datos centrales.

### Datos agrupados

Si se tienen  $k$  intervalos expresados  $[l_0, l_1), \dots, [l_{k-1}, l_k]$ , siendo  $l_0$  el mínimo,  $l_k$  el máximo y  $l_1 = l_0 + A$ . Para calcular la mediana se realiza los siguientes procedimientos; primero determinamos en la columna de frecuencias relativas acumuladas ( $H_i$ ), el valor correspondiente al 50% de los datos, para observar en que intervalo tiene una frecuencia absoluta acumulada ( $H_i$ ) que alcance o sobrepase el 50%, es decir que cumpla que  $H_{i-1} < 50\% \leq H_i$ , en donde este será el intervalo mediano o intervalo donde está localizada la mediana. Una vez identificado el intervalo se utiliza la siguiente fórmula.

$$Me = l_{i-1} + \frac{\frac{n}{2} - F_{i-1}}{f_i} \cdot (l_i - l_{i-1}) \quad (2-17)$$

Siendo  $l_{i-1}$  el extremo inferior del intervalo donde se encuentra la Mediana,  $F_{i-1}$  la frecuencia acumulada inmediatamente anterior al intervalo mediano,  $f_i$  la frecuencia absoluta del intervalo mediano y  $(l_i - l_{i-1})$  la amplitud del intervalo que contiene a la Mediana. (Chaparro, et al., 2018)

### 2.2.3.3. Moda ( $Mo$ )

#### Datos sin agrupar

La moda como su nombre lo indica, se refiere al valor más común o el de mayor frecuencia en un conjunto de datos. Una distribución de datos puede tener una moda, dos modas (bimodal) o varias modas (multimodal)

Sosa Martínez, Ospina Forero, & Berdugo Camacho (2013) proponen la siguiente definición “se llama **moda** a cualquier valor que maximize la distribución de frecuencias de un conjunto de datos”

### Datos agrupados

Cuando los datos están agrupados en una tabla de frecuencias por intervalos se determina la moda de la siguiente manera; primero se determina en la columna de las frecuencias absolutas  $f_i$  el intervalo que tenga la mayor frecuencia. Una vez identificado el intervalo modal o intervalo que contiene la moda, se aplica la siguiente fórmula.

$$Mo = l_{i-1} + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot (l_i - l_{i-1}) \quad (2-18)$$

Siendo  $l_{i-1}$  el extremo inferior del intervalo donde se encuentra la Moda,  $f_{i-1}$  la frecuencia absoluta inmediatamente anterior,  $f_{i+1}$  la frecuencia absoluta inmediatamente posterior al intervalo modal,  $f_i$  la frecuencia absoluta del intervalo mediano y  $(l_i - l_{i-1})$  la amplitud del intervalo que contiene a la Moda. Para destacar, la moda no puede ser única y no tiene propiedades aritméticas directas y fáciles de aplicar como la media aritmética.

### 2.2.4. Medidas de posición

Las medidas de posición permiten conocer otras características de la distribución de los datos que permiten dividir la información a conveniencia. Se define el percentil de orden  $k$ , como la observación  $P_k$ , que deja por debajo de si el  $k\%$  de la muestra. Esta definición es similar a la de mediana, pues ya que como efecto de la definición, se puede concluir que  $Me = P_{50}$ . El percentil se ubica en cualquier lugar de la distribución, por lo tanto, no se puede considerar como una medida de tendencia central, sino más bien como una medida de posición. *“El  $p$ -ésimo percentil es un valor tal que por lo menos un  $p\%$  de los datos tienen máximo dicho valor y al menos un  $(100 - p)\%$  de los datos tienen es valor o uno mayor.”* (Sosa Martínez, et al., 2013)

Los cuartiles, denominados  $Q_k$ , son un caso particular de los percentiles, y se definen como primer cuartil  $Q_1 = P_{25}$ ; segundo cuartil  $Q_2 = P_{50}$ , que es el equivalente a la mediana y el tercer cuartil  $Q_3 = P_{75}$ . De forma análoga se definen los deciles, para este caso se divide el conjunto de datos en 10 grupos de igual tamaño  $D_1, \dots, D_8, D_9$ . Es decir  $D_i = P_{10 \cdot i}$  donde  $i = 1, 2, 3, \dots, 9$

**Datos sin agrupar**

Para calcular el  $p$ -ésimo percentil de un conjunto de  $n$  datos se realiza el siguiente procedimiento; primero se ordenan los datos de forma ascendente, posteriormente se calcula el índice  $i$

$$i = \frac{k n}{100} \quad (2-19)$$

Donde  $k$  es el percentil de interés y  $n$  el tamaño de la muestra. Una vez obtenido el índice  $i$  de la muestra, se procede a determinar el percentil de acuerdo a uno de los siguientes casos. (Sosa Martínez, et al., 2013, p. 57)

- i. Si el índice  $i$  no es entero, entonces se hace el redondeo al entero siguiente; este valor aproximado indica la posición del percentil que se consideró.
- ii. Si el índice  $i$  es entero, entonces se realiza el promedio de los valores ubicados en las posiciones  $i$  y  $i + 1$

**Datos agrupados**

Cuando los datos se encuentran agrupados en una tabla de frecuencias por intervalos el procedimiento se realiza igual que el de la mediana para datos agrupados. (2.2.3.2) Una vez identificado el intervalo que contiene el  $p$ -ésimo percentil se aplica la siguiente fórmula.

$$P_k = l_{i-1} + \frac{\frac{k n}{100} - F_{i-1}}{f_i} \cdot (l_i - l_{i-1}) \quad (2-20)$$

**2.2.5. Medidas de dispersión.**

Cuando se realiza el análisis de información no bastan las medidas de tendencia central y de posición, debido a que se puede llegar a conclusiones erróneas, dado que pueden existir datos atípicos. Los estadísticos de dispersión indican si las observaciones están próximas entre sí o si por el contrario están muy dispersas.

### 2.2.5.1. El rango

Esta es una medida de dispersión que se utiliza en la construcción de las tablas de distribución de frecuencias de datos agrupados (ver sección 2.2.2.2) y en la construcción de histogramas.

Dado  $x_1, x_2, x_3, x_4, \dots, x_n$  un conjunto de  $n$  datos, el rango se calcula como la diferencia entre el valor máximo y mínimo.

$$R_X = x_{max} - x_{min} \quad (2-10)$$

### 2.2.5.2. El rango intercuartílico

Se denota  $RI_X$ , es un valor que no está influenciado por los valores extremos de los datos como el rango. En este caso el rango intercuartílico se concentra en un 50% de los datos que no hacen parte del 25% de los valores más bajos ni del 25% de los valores más altos. Dado  $x_1, x_2, x_3, x_4, \dots, x_n$  un conjunto de  $n$  datos de una variable  $X$ , el rango intercuartílico se calcula como la diferencia entre los valores del tercer cuartil ( $Q_3$ ) y primer cuartil ( $Q_1$ ) del conjunto de datos.

$$RI_X = Q_3 - Q_1 \quad (2-21)$$

Al respecto Sosa Martínez, Ospina Forero, & Berdugo Camacho (2013), “*el rango es una medida asociada con la extensión de todo conjunto de datos y el rango intercuartílico es una medida relacionada con la amplitud correspondiente al 50% de los datos intermedios, ninguna de estas medidas de dispersión tiene en cuenta cómo se distribuyen las observaciones de la variable de estudio respecto a alguna medida de tendencia central, como el promedio,*”

Una medida que toma en cuenta tal repartición de los datos respecto a la media aritmética es la varianza. Ya que esta medida evalúa la manera en que fluctúan los valores de una variable con respecto al promedio



### 2.2.5.3. La varianza y la desviación estándar<sup>1</sup>

#### Datos sin agrupar

Dados  $x_1, x_2, x_3, x_4, \dots, x_n$  un conjunto de  $n$  datos de una variable  $X$ . La varianza muestral de este conjunto de datos que están organizados en una tabla de distribución, se denota  $S^2$  y está dado por.

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (2-22)$$

El resultado de la varianza está expresado en unidades cuadradas, este aspecto puede dificultar la interpretación ya que se debe relacionar unidades cuadráticas con unidades unidimensionales y se puede llegar a conclusiones incorrectas. La parte positiva de su raíz cuadrada. La desviación estándar o desviación típica, resuelve esta dificultad.

Donde  $\bar{x}$  corresponde a la media aritmética. La desviación estándar se define como la raíz cuadrada positiva de la varianza, es decir.

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (2-23)$$

#### Datos agrupados

En el caso de los datos agrupados en intervalos, el cálculo de la varianza y la desviación estándar la fórmula tiene un cambio, ya que ahora se considerará la marca de clase en vez del  $x_i$ . Es decir

$$S^2 = \frac{\sum_{i=1}^k (X'_i - \bar{x})^2 \cdot f_i}{n - 1} \quad (2-24)$$

<sup>1</sup> Se tomará para este caso únicamente la varianza y la desviación estándar muestral ya que son los conceptos que se indagan en la básica secundaria con este grupo de estudiantes.

Y la desviación estándar estaría dada por

$$S = \sqrt{\frac{\sum_{i=1}^k (X'_i - \bar{x})^2 \cdot f_i}{n - 1}} \quad (2-25)$$

Donde  $X'_i$  indica la marca de clase del i-ésimo intervalo

## 2.2.6. Gráficos estadísticos

En la construcción de un gráfico estadístico se deben considerar aspectos fundamentales como lo son, el tipo de variable y las características de estudio ya que el principal objetivo de estas figuras es mostrar de una manera muy clara el comportamiento de una o varias variables. (Sosa Martínez et al., 2013, p.31) A continuación haremos una breve descripción de algunos gráficos que se consideraron en el desarrollo de este trabajo, tomaremos como base el trabajo realizado por (Rojas et al., 2017)

### 2.2.6.1. Gráficos para variables cualitativas

En este caso hablaremos de un tipo de gráfico que se conoce como el **diagrama de barras**, en este caso se utilizan dos ejes perpendiculares de modo que en alguno de ellos se representan las diferentes modalidades de la variable que se representan con una barra. En el otro eje se representan las frecuencias absolutas o relativas sobre una recta numérica que inicia desde cero y toma valores positivos. En el eje donde está ubicada cada una de las categorías, se levantan las barras con una altura equivalente a la respectiva frecuencia absoluta o relativa. (Rojas, et al., 2017 p. 14)

El **diagrama de Pareto** es un gráfico de barras utilizado exclusivamente para variables de escala nominal, el objetivo es identificar los aspectos prioritarios que hay que tratar en un estudio (con respecto a las modalidades de la variable), así como sus causas más importantes. Para la construcción se elabora una tabla de distribución de frecuencias. (ver sección 2.2.2.1) Luego, se construye un gráfico de barras, ordenando las categorías de mayor a menor frecuencia, de izquierda a derecha.

### 2.2.6.2. Gráficos para variables cuantitativas

#### Gráficos para variables continuas

**Histograma.** Se construye teniendo en cuenta la tabla de distribución de frecuencias que aparece en la sección 2.2.2.2. Se utilizan barras en sentido vertical de manera que a cada marca de clase le corresponde una barra cuya base es el intervalo de la marca de clase correspondiente. La altura de cada barra es el equivalente a la frecuencia absoluta  $f_i$  del intervalo,  $i = 1, \dots, k$  ( $k$  clases).

**Polígono de frecuencias.** Su construcción se puede realizar junto con la del histograma teniendo en cuenta los rectángulos ya elaborados en el histograma, donde se localizan los puntos medios en la parte superior del rectángulo y luego unir mediante líneas rectas dichos puntos. Es decir, se tienen  $k$  puntos de coordenadas  $(X'_i, f_i)$  y se unen con líneas rectas, donde  $f_i$  representa la frecuencia absoluta y  $X'_i$  representa la marca de clase, que es el punto medio de cada intervalo,  $i = 1, \dots, k$

#### Diagrama de caja y bigotes (Boxplot)

Es un gráfico que sirve para identificar algunas características fundamentales de una distribución de un conjunto de datos como lo son, la dispersión, los datos atípicos, la localización, para la construcción de este diagrama se necesitan calcular los cuartiles (ver sección 2.2.4), el cuartil 1, cuartil 3, el rango intercuartílico (ver sección 2.2.5.2) y los límites que definiremos a continuación.

Dado  $x_1, x_2, x_3, x_4, \dots, x_n$  un conjunto de  $n$  datos de una variable  $X$ , los límites inferior y superior están dados por.

$$LI = Q_1 - 1.5 \cdot RI_x \quad (2-26)$$

$$LS = Q_3 + 1.5 \cdot RI_x \quad (2-27)$$

Para realizar la construcción vertical de este gráfico se realizan los siguientes pasos, según Sosa Martínez et al., (2013 p. 68)

- i. Obtener los cuartiles  $Q_1$ ,  $Q_2$  y  $Q_3$  del conjunto de datos con  $n$  elementos.
- ii. Dibujar un rectángulo paralelo al eje  $y$  de tal manera que su largo este determinado por  $Q_1$  y  $Q_3$ . El ancho de la caja no es de relativa importancia, pero debe ser menor que el largo del rectángulo.
- iii. Trazar los segmentos de la caja así:
  - a. Si no hay datos atípicos, desde el mínimo hasta  $Q_1$  y desde  $Q_3$  hasta el máximo.
  - b. Si hay datos atípicos, hasta  $LI$  y/o  $LS$ , dependiendo del caso atípico.
- iv. Señalar y resaltar los valores que se encuentran por fuera del intervalo  $[LI, LS]$ .
- v. Trazar una línea horizontal dentro de la caja a la altura de  $Q_2$ .

A continuación en la Figura 2-1, se muestran dos ejemplos de boxplot con datos atípicos y sin datos atípicos tomados de (Rojas, 2017).

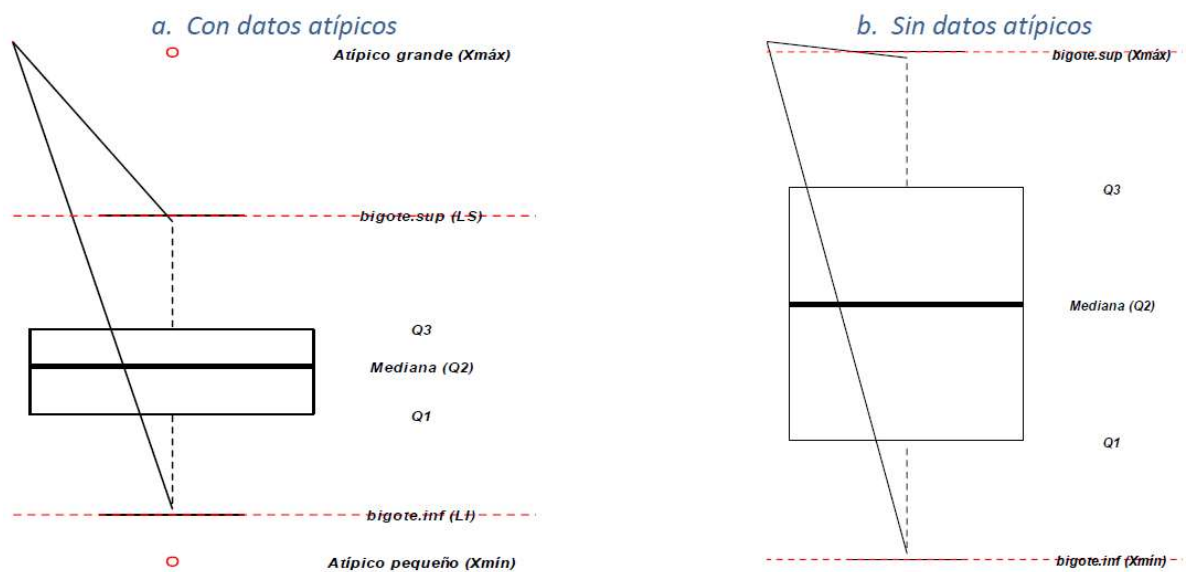


Figura 2-1 Ejemplos de diagrama de cajas y bigotes. (Boxplot)

### Gráficos para variables discretas

Es posible utilizar gráficos de barras para este tipo de variables, por medio de un **diagrama diferencial**<sup>2</sup> donde las barras deben ser estrechas para indicar que se tratan de valores discretos, aunque este diagrama es utilizado de esta manera siempre y cuando el número de categorías sea pequeño.

## 2.3. Marco metodológico: Investigación – Acción

Al revisar la evolución de la investigación esta se encuentra dividida en etapas que se encuentran bien demarcadas por dos grandes enfoques que han conducido al desarrollo de diferentes metodologías que permiten entrar en el campo de la investigación por diferentes vías. Estos dos grandes enfoques son distinguidos con los términos de cuantitativa y cualitativa y albergan un abanico de alternativas metodológicas en donde se encuentra la investigación-acción con sus diferentes connotaciones que les dan los distintos autores e investigaciones. (Colmenares & Piñero, 2008)

El comienzo de la investigación-acción se da hacia la década de los 40 y se sitúa en los trabajos llevados a cabo por el psicólogo estadounidense de origen alemán Kurt Lewin a raíz de la segunda guerra mundial y por petición de la administración del gobierno, él realiza estudios sobre la modificación de los hábitos alimenticios ante la escasez de determinados artículos, el objetivo de estos trabajos era resolver problemas prácticos y urgentes, adoptando los investigadores el papel de agentes de cambio, en colaboración directa con aquellas personas a quienes iban destinadas las propuestas de intervención. Aquí ya se vislumbran algunos de los rasgos característicos de la investigación-acción (el conocimiento, la intervención, la mejora, la colaboración) (Suárez , 2002)

Colmenares y Piñero (2008) también hacen mención a que la investigación-acción en su transitar histórico ha desarrollado dos grandes tendencias o vertientes: una tendencia netamente sociológica, cuyo punto de partida fueron los trabajos de Kurt Lewin y la otra

---

<sup>2</sup> Es un diagrama donde se representa las frecuencias absolutas o relativas.

vertiente es más educativa, y está inspirada en las ideas de Paulo Freire (1974) en Brasil, L. Stenhouse (1988) y Jhon Elliott (1981, 1990) discípulo de Stenhouse en Inglaterra, así como por Carr y Stephen Kemmis (1988) de la Universidad de Deakin en Australia.

Mirando un ámbito más educativo, en palabras de Suarez (2002) *“la investigación-acción es una forma de estudiar, de explorar, una situación social, en nuestro caso educativa, con la finalidad de mejorarla en la que se implican como “indagadores” los implicados en la realidad investigada”* y para profundizar en esta modalidad propone cuatro preguntas claves para el desarrollo ¿Qué se investiga? ¿Quién? ¿Cómo? y ¿para qué?

La investigación-acción es una metodología hacia el cambio de educativo y se caracteriza por ser un proceso que como señalan Kemmis y MacTaggart citado en (Bausela, 2004) 1. Se construye desde y para la práctica. 2. Pretende mejorar la práctica a través de su transformación, al mismo tiempo que procura comprenderla. 3. Demanda la participación de los sujetos en la mejora de sus propias prácticas. 4. Exige una actuación grupal por la que los sujetos implicados colaboran coordinadamente en todas las fases del proceso de investigación. 5. Implica la realización de análisis crítico de las situaciones. 6. Se configura como una espiral de ciclos de planificación, acción, observación y reflexión.

Autores como (Garcia-Carmona, 2009) establecen que cuando la investigación-acción está en un ámbito educativo prevalece una metodología cuantitativa, esto se debe a que la investigación centra su atención en aspectos relativos a la distribución del aula, las relaciones entre el profesor con los estudiantes. En general, en los aspectos locales de los acontecimientos que tiene lugar en los grupos de la clase. No obstante, en la investigación acción, también debe estar presente la metodología cuantitativa. Su uso se es notable a través del empleo de la Estadística Descriptiva.

El carácter del trabajo que se desarrolló es mixto porque integra los enfoques cuantitativo y cualitativo. Por un lado, en el enfoque cualitativo se realiza un seguimiento a los procesos académicos realizados por los estudiantes en la interpretación y análisis de medidas descriptivas y en la utilización de los gráficos descritos en la sección 2.2.6 en cuanto los aspectos cuantitativos por medio de la comparación entre los conocimientos previos del estudiante y los resultados alcanzados al finalizar el proyecto.

Las variables de estudio en esta investigación están asociadas a la lectura, interpretación y análisis de las medidas descriptivas por parte de los estudiantes como también a la construcción de los gráficos estadísticos anteriormente descritos.

## **2.4. Marco didáctico: Trabajo por proyectos.**

Una de las estrategias didácticas de enseñanza aprendizaje en la estadística es el trabajo con proyectos, cuya principal finalidad es involucrar el conocimiento estadístico en la resolución de problemas de índole disciplinar o social. En el desarrollo de los trabajos por proyectos en el aula, se busca contextualizar la estadística y buscar el interés y la motivación por parte de los estudiantes, para que desarrollen un problema de investigación. De tal manera que durante el desarrollo del proyecto los estudiantes generen competencias matemáticas, como las propuestas por el MEN para el área de matemáticas, es decir, competencias como la comunicación, el razonar, la resolución de problemas, etc. Con el fin de que los estudiantes mejoren sus procesos argumentativos y sean propositivos en cuanto a la toma de decisiones de alguna temática específica.

Al respecto (Batanero & Díaz , Estadística con proyectos, 2011) manifiestan que la idea del trabajo con proyectos es fomentar el razonamiento estadístico, más que el aprendizaje rutinario y descontextualizado de conceptos y propiedades. Por ejemplo, haciendo uso de tecnología se puede aplicar la estadística con gran facilidad, lo que hace que cobre mayor importancia las actividades interpretativas que las de cálculo rutinario.

Autoras como Watson citado en (Batanero, Díaz, Contreras, & Arteaga, 2011) establece que es muy importante que una persona adulta estadísticamente culta tenga las siguientes habilidades, las cuales ella las define en una jerarquía de niveles de cultura estadística útil para evaluar la comprensión de los estudiantes, los niveles que define son los siguientes:

1. El desarrollo del conocimiento básico de los conceptos estadísticos y probabilísticos.
2. La comprensión de los razonamientos y argumentos estadísticos cuando se presentan dentro de un contexto más amplio de algún informe en los medios de comunicación o en el trabajo.
3. Una actitud crítica que se asume al cuestionar argumentos que estén basados en evidencia estadística.

Al respecto Batanero y Díaz (2011, p.13) manifiestan que *“la mejor forma de seguir las anteriores recomendaciones es introducir en las clases de estadística el trabajo con proyectos, algunos de los cuales son planteados por el profesor y otros escogidos libremente por los alumnos. En lugar de introducir los conceptos y técnicas descontextualizadas, o aplicadas únicamente a problemas tipo, difíciles de encontrar en la vida real, se trata de presentar las diferentes fases de una investigación estadística: planteamiento de un problema, decisión sobre los datos a recoger, recogida y análisis de datos y obtención de conclusiones sobre el problema planteado.”*

Las fases iniciales del trabajo con proyectos en primera instancia es elegir la pregunta de investigación que debe ser realizada por el estudiante y orientada por el docente para dar paso al problema, teniendo en cuenta los siguientes aspectos: *“¿Qué quieres probar?, ¿Qué tienes que medir?, ¿Qué datos necesitas?, ¿Cómo encontrarás tus datos?, ¿Qué harás con ellos?, ¿Crees que puedes hacerlo?, ¿Podrás contestar tu pregunta? [y] ¿Para qué te servirán los resultados?”* (Batanero y Díaz, 2011, p. 23).

Cuando los estudiantes estén en la realización de sus proyectos, es necesario que los estudiantes elaboren los informes acerca de su “investigación” de forma clara y lógica, porque además se está motivando al desarrollo de la competencia argumentativa ya que los estudiantes deben relatar sus decisiones, acciones e interpretaciones. A su vez, para el proceso de evaluación del proyecto es importante estar realizando los procesos de retroalimentación constante para mejorar los aprendizajes y por ende las competencias, además se deben en cuenta los distintos aspectos del conocimiento matemático como los del conocimiento procedimental. Además de evaluar aspectos del conocimiento matemático y procedimental también se debe tener en cuenta la relevancia de la pregunta, el diseño de la investigación con énfasis en la población y la muestra, el análisis de datos con respecto a la pregunta de investigación, las conclusiones referentes al análisis, la reflexión del proceso realizado, la presentación de los resultados con énfasis en la claridad de los gráficos y la creatividad, y la originalidad del estudiante. (Rojas, 2017)



### 3. Estrategia didáctica

Para reforzar el aprendizaje de los estudiantes, los docentes deben ser eficaces en comprobar constantemente la comprensión que éstos están logrando. Además, tienen que darles a conocer la importancia de que ellos mismos asuman la responsabilidad de reflexionar sobre su propio progreso en el aprendizaje. (Shepard, 2006)

Shepard también manifiesta en su modelo de evaluación formativa un aspecto que se denomina ***evaluación del conocimiento previo*** del cual establece que es esencial para el aprendizaje, ya que el proceso de aprendizaje puede concebirse como lo que hacemos para conectar e integrar una nueva comprensión con el conocimiento existente.

El desarrollo de la propuesta inicia con la elaboración de una prueba diagnóstica que al aplicarla arrojará información acerca de los conocimientos previos y de las fortalezas o debilidades en los conceptos que se esperaba tengan los estudiantes de grado décimo según los estándares básicos de competencias establecidos por el MEN. De acuerdo a la información que se recopiló de la prueba y con los diálogos realizados con los estudiantes sobre las preguntas de la prueba, se estructuró un proyecto con preguntas dirigidas hacia la interpretación y análisis de las medidas descriptivas como también la de los gráficos estadísticos descritos en la sección 2.2.6.

Después del trabajo por proyecto se implementó nuevamente esta prueba con algunas modificaciones, para identificar el efecto que tuvo la estrategia didáctica (Trabajo con proyectos) con respecto a los conocimientos estadísticos adquiridos por los estudiantes.

### 3.1. Prueba diagnóstica

La estructura de la prueba diagnóstica se basó en los estándares académicos del MEN para el grado décimo. Para determinar las fortalezas o debilidades de los estudiantes se organizó una prueba de 10 preguntas de las cuales 7 fueron tomadas del trabajo de Chaparro (2018); estas 7 preguntas seleccionadas hacen parte de las preguntas liberadas por el ICFES del cuadernillo pruebas saber para grado noveno del año 2013 y del libro Estocástica y su Didáctica para Maestros (Batanero & Godino, 2002). Las otras 3 preguntas, 2 fueron tomadas de la Guía N°1 del curso de estadística de la Universidad de Atacama de la Facultad de Ingeniería del profesor Salinas (2012) y la otra surgió de las revisiones realizadas con el director del presente trabajo; cabe aclarar que las preguntas anteriormente mencionadas también fueron discutidas y adaptadas con el profesor Luis F. Grajales director del trabajo.

La prueba se distribuye de la siguiente manera, las preguntas de la 1 a la 9 son preguntas de selección múltiple con única respuesta, cada una de ellas con cuatro opciones de respuesta. La pregunta 10 está conformada por dos partes, a y b, en donde los estudiantes realizan los cálculos de algunas medidas de tendencia central, de dispersión y posición para datos enumerados de conjunto de datos establecido. (Ver Anexo: A)

#### 3.1.1. Análisis a priori de la prueba diagnóstica

El análisis conceptual y procedimental se fundamenta en los Lineamientos Curriculares y los Estándares Básicos en Competencias en Matemáticas y se realizará con el esquema planteado por el ICFES en las cartillas Saber 9° Lineamientos para las aplicaciones muestral y censal 2016 (ICFES, 2016), ver **Tabla 3.1**

**Tabla 3.1** Esquema propuesto por el ICFES para el análisis conceptual y procedimental

Pregunta ____	
Competencia	
Componente	
Afirmación	
Nivel desempeño	
Respuesta correcta	
Breve descripción de la respuesta correcta	

En la primera fila de la **Tabla 3.1** se coloca la pregunta realizada, en la segunda fila aparece la casilla de la “competencia” que hace referencia a los cinco procesos generales de la actividad matemática que se contemplan en los Lineamientos Curriculares de Matemáticas (1998, p.51), aunque para este caso el ICFES las reúne en tres grupos. La tercera fila aparece la casilla de la “componente” que los Lineamientos Curriculares (et al., 1998 p.37) llaman conocimientos básicos que describe los cinco pensamientos, (Numérico, variacional, Geométrico, métrico y aleatorio o estocástico) en este caso el ICFES también los reorganiza y los distribuye en tres grupos. En la cuarta fila, casilla de “afirmación” se describen enunciados que serán elaborados de acuerdo con lo establecido en los Estándares Básicos de Competencias en Matemáticas. En la quinta fila se encuentra la casilla “Nivel de desempeño” en donde se asigna un puntaje de acuerdo a unos niveles de desempeño que se establecieron con el objetivo de complementar el puntaje numérico que se otorga a los estudiantes. Consisten en una descripción cualitativa de las habilidades y conocimientos que podrían tener si se ubican en determinado nivel. Pero para nuestro caso no tendremos en cuenta esta casilla ya que no tiene trascendencia en el análisis que se quiere realizar, así que suprimiremos esta casilla. Finalmente, en las últimas dos casillas aparecerá la respuesta correcta y luego se hará una breve descripción de la respuesta correcta.

**Tabla 3.2** Análisis conceptual y procedimental pregunta 1.

<b>PREGUNTA 1</b>	
<b>Competencia</b>	Resolución de problemas
<b>Componente</b>	Aleatorio
<b>Afirmación</b>	( <b>Estándar 6° a 7°</b> ) Resuelve y formula problemas a partir de un conjunto de datos presentados en tablas, diagramas de barras y diagrama circular
<b>Respuesta correcta</b>	C
<b>Breve descripción de la respuesta correcta:</b> <b>Una vez identificadas las frecuencias correspondientes a cada uno de los valores que tomó la variable “número de hermanos”, se realiza la suma de las frecuencias de cada una de las variables. La suma de las frecuencias es igual a 58.</b>	

**Tabla 3.3** Análisis conceptual y procedimental pregunta 2.

<b>PREGUNTA 2</b>	
<b>Competencia</b>	Resolución de problemas
<b>Componente</b>	Aleatorio
<b>Afirmación</b>	(Estándar 6° a 7°) Usa medidas de tendencia central (media, mediana, moda) para interpretar el comportamiento de un conjunto de datos.
<b>Respuesta correcta</b>	A
<b>Breve descripción de la respuesta correcta:</b> Una vez identificado las frecuencias correspondientes a cada uno de los valores que tomó la variable “número de hermanos” se procede a calcular el promedio $\bar{x} = \frac{0 \times 16 + 1 \times 25 + 2 \times 12 + 3 \times 4 + 4 \times 1}{58} = 1.12$	

**Tabla 3.4** Análisis conceptual y procedimental pregunta 3.

<b>PREGUNTA 3</b>	
<b>Competencia</b>	Resolución de problemas
<b>Componente</b>	Aleatorio
<b>Afirmación</b>	(Estándar 6° a 7°) Usa medidas de tendencia central (media, mediana, moda) para interpretar el comportamiento de un conjunto de datos.
<b>Respuesta correcta</b>	D
<b>Breve descripción de la respuesta correcta:</b> Una vez identificado las frecuencias correspondientes a cada uno de los valores que tomó la variable “número de hermanos” se determina cuál es el más común o el de mayor frecuencia. Para este caso la moda es 1.	

**Tabla 3.5** Análisis conceptual y procedimental pregunta 4.

<b>PREGUNTA 4</b>	
<b>Competencia</b>	Comunicación
<b>Componente</b>	Aleatorio

<b>Afirmación</b>	<b>(Estándar 10° a 11°)</b> Interpreto y comparo resultados de estudios con información estadística provenientes de medios de comunicación.
<b>Respuesta correcta</b>	D
<b>Breve descripción de la respuesta correcta:</b> Identificar en el histograma los extremos inferior y superior de los siete intervalos de la variable “altura” y sus respectivas frecuencias correspondientes a la altura de los rectángulos. Para este caso se debe identificar el intervalo que contenga la altura 160 cm. Luego, observar y sumar las frecuencias de los rectángulos que estén a su derecha, incluyendo la frecuencia del rectángulo de intervalo [160, 165) . Una vez totalizado el valor corresponde a 34 personas.	

**Tabla 3.6** Análisis conceptual y procedimental pregunta 5

<b>PREGUNTA 5</b>	
<b>Competencia</b>	Comunicación
<b>Componente</b>	Aleatorio
<b>Afirmación</b>	<b>(Estándar 10° a 11°)</b> Interpreto y comparo resultados de estudios con información estadística provenientes de medios de comunicación.
<b>Respuesta correcta</b>	B
<b>Breve descripción de la respuesta correcta:</b> Se debe Identificar en el histograma los extremos inferior y superior de los siete intervalos de la variable “altura” y sus respectivas frecuencias correspondientes a la altura de los rectángulos. Para este caso se debe identificar el intervalo que no contenga la altura 160 cm. Luego, observar y sumar las frecuencias de los rectángulos que estén a su izquierda, incluyendo la frecuencia del rectángulo de intervalo [155, 160) . Una vez totalizado el valor corresponde a 24 personas.	

**Tabla 3.7** Análisis conceptual y procedimental pregunta 6

<b>PREGUNTA 6</b>	
<b>Competencia</b>	Comunicación
<b>Componente</b>	Aleatorio

<b>Afirmación</b>	<b>(Estándar 8° a 9°)</b> Reconozco cómo diferentes maneras de presentación de información pueden originar distintas interpretaciones.
<b>Respuesta correcta</b>	D
<b>Breve descripción de la respuesta correcta:</b> En la tabla de distribución de frecuencias la primera columna corresponde a la variable “edad” distribuida en seis intervalos, a cada uno de ellos le corresponde una frecuencia relativa que multiplicada por 100, se obtiene el porcentaje de cada una con respecto a la población. Dado que se desea saber el número de empleados que tienen edad mayor o igual que 21 años y menor de 25 años, identificamos y sumamos las respectivas frecuencias relativas correspondientes a los intervalos que estén en el rango establecido. Esto corresponde a 116 empleados de los 400 empleados que tiene la empresa.	

**Tabla 3.8** Análisis conceptual y procedimental pregunta 7

<b>PREGUNTA 7</b>	
<b>Competencia</b>	Comunicación
<b>Componente</b>	Aleatorio
<b>Afirmación</b>	<b>(Estándar 8° a 9°)</b> Reconozco cómo diferentes maneras de presentación de información pueden originar distintas interpretaciones.
<b>Respuesta correcta</b>	B
<b>Breve descripción de la respuesta correcta:</b> En la tabla de distribución de frecuencias la primera columna corresponde a la variable “edad” distribuida en seis intervalos, a cada uno de ellos le corresponde una frecuencia relativa que multiplicada por 100, se obtiene el porcentaje de cada una con respecto a la población. Dado que se desea saber el porcentaje de los empleados que tienen 25 años o más. Entonces se suman las frecuencias relativas de los intervalos que inicien en el intervalo [155,160) hasta el último intervalo $0.28 + 0.24 + 0.14 = 0.66$ . Es decir el 66%.	

**Tabla 3.9** Análisis conceptual y procedimental pregunta 8

<b>PREGUNTA 8</b>
-------------------

<b>Competencia</b>	Resolución de problemas
<b>Componente</b>	Aleatorio
<b>Afirmación</b>	<b>(Estándar 6° a 7°)</b> Usa medidas de tendencia central (media, mediana, moda) para interpretar el comportamiento de un conjunto de datos.
<b>Respuesta correcta</b>	C
<b>Breve descripción de la respuesta correcta:</b> La gráfica muestra los periodos facturados en bimestres con respecto al respectivo consumo de agua en $m^3$ . Para responder la pregunta se debe promediar el consumo agua de la vivienda durante los 4 bimestres, es decir $\bar{x} = \frac{14+13+18+19}{4} = 16 m^3$	

**Tabla 3.10** Análisis conceptual y procedimental pregunta 9

<b>PREGUNTA 9</b>	
<b>Competencia</b>	Comunicación
<b>Componente</b>	Aleatorio
<b>Afirmación</b>	<b>(Estándar 10° a 11°)</b> Interpreto y comparo resultados de estudios con información estadística provenientes de medios de comunicación.
<b>Respuesta correcta</b>	A
<b>Breve descripción de la respuesta correcta:</b> El boxplot muestra los valores de los cuartiles, para solucionar la pregunta es necesario realizar la interpretación de las medidas de posición, en este caso la medida correspondiente al $Q_1$ , es decir, que acumula el 25% de los datos.	

**Tabla 3.11** Análisis conceptual y procedimental pregunta 10 (a)

<b>PREGUNTA 10 (a)</b>	
<b>Competencia</b>	Planteamiento y Resolución de problemas <sup>3</sup>
<b>Componente</b>	Aleatorio

<sup>3</sup> En esta competencia también están inmersas la elaboración, comparación y ejercitación de procedimientos que se encuentran en los Lineamientos Curriculares de Matemáticas.

<b>Afirmación</b>	<b>(Estándar 6° a 7°)</b> Usa medidas de tendencia central (media, mediana, moda) para interpretar el comportamiento de un conjunto de datos.
<b>Respuesta correcta</b>	$\bar{x} = 14.15$ $Me = 13.7$
<b>Breve descripción de la respuesta correcta:</b> <b>Para este caso se deben calcular las medidas de tendencia central establecidas, media <math>\bar{x}</math> y la mediana del conjunto de datos dado.</b>	

**Tabla 3.12** Análisis conceptual y procedimental pregunta 10 (b)

<b>PREGUNTA 10 (b)</b>	
<b>Competencia</b>	Planteamiento y Resolución de problemas
<b>Componente</b>	Aleatorio
<b>Afirmación</b>	<b>(Estándar 6° a 7°)</b> Usa medidas de tendencia central (media, mediana, moda) para interpretar el comportamiento de un conjunto de datos.
<b>Respuesta correcta</b>	$Q_1 = 12.3$ $Q_3 = 15.3$ $S = 2.64$
<b>Breve descripción de la respuesta correcta:</b> <b>Para este caso se deben calcular las medidas de posición y dispersión establecidas, media <math>Q_1</math>, <math>Q_3</math> y <math>S</math> del conjunto de datos dado</b>	

### 3.1.2. Descripción general de las sesiones previas al desarrollo del proyecto.

Una vez obtenidos los resultados de la prueba diagnóstica y con las preguntas manifestadas por los estudiantes durante el desarrollo de la prueba se realizaron 4 sesiones con una duración de 90 minutos cada una, donde se abordaron algunas de las temáticas descritas en la sección 2.2. En el siguiente capítulo se presentará una descripción más detallada de los resultados de la prueba diagnóstica.



Para el desarrollo de estas secciones se utilizaron dos bases de datos obtenidas del DANE, una correspondía al Censo Habitantes de la calle en Bogotá D.C. en el año 2017, y la otra base correspondía a las Estadísticas Vitales de Colombia año 2017. De ambas bases se seleccionaron las sub-bases de 76 y 106 datos respectivamente, junto con las variables que sirvieran para realizar el abordaje de las temáticas en las que se identificaron dificultad, el proceso de la selección de datos y variables fue el resultado de las asesorías con el director del presente trabajo. A continuación en la Tabla 3.13 y la Tabla 3.14 se muestran algunas de las variables seleccionadas para el desarrollo de las sesiones. (Ver Anexo: Hojas de códigos. Censo Habitantes de calle Bogotá año 2017, Estadísticas vitales de Colombia año 2017 y Estadísticas de la prueba Saber 11 año 2017)

Tabla 3.13 Datos de la encuesta Habitantes de calle Bogotá D.C. año 2017

V1	V2	V3	V4	V5	V6	V7	V8	V9
Habitante	Localidad	Lugar de la entrevista	Años cumplidos	Género	Principal razón HC	Leer y Escribir	Nivel educativo	Consigue dinero
1	18	1	37	1	11	1	3	5
2	18	1	25	1	5	1	3	5
3	18	1	43	1	1	1	2	5
:	:	:	:	:	:	:	:	:

Tabla 3.14 Datos de Estadísticas vitales Colombia año 2017.

V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V11
NEONATO	COD_D TO	SEXO	PESO_NAC	TALLA_NAC	MES DE NACIMIENTO	ATENDIDO	TIPO_PARTO	EDAD_MADRE	...	NIV_EDUP
1	11	2	5	5	1	1	1	1	...	99
2	18	2	5	4	1	1	2	1	...	99
3	52	2	5	4	1	1	2	6	...	4
:	:	:	:	:	:	:	:	:	:	:

Para el desarrollo de las sesiones los estudiantes se organizaron en grupos de máximo 5 personas, a cada grupo se le asignó una base de datos y con éstas se desarrollaron las actividades propuestas para cada sesión. La primera sesión, el objetivo era que los estudiantes identificaran los tipos de variables estadísticas y que realizaran la construcción de tablas de frecuencia de datos agrupados y no agrupados para un tipo de variable en particular. Para esto primero se les solicitó a los estudiantes que determinaran la naturaleza de un listado de variables. Posteriormente como segunda parte de la actividad, se les solicitó que escogieran una variable cuantitativa continua y que realizaran una tabla de frecuencias de datos agrupados.

En la segunda sesión se abordaron los temas de medidas de tendencia central, el objetivo era identificar para qué tipo de variables es posible calcular este tipo de medidas y determinar el valor de cada una de ellas realizando las respectivas interpretaciones para cada caso. A cada grupo se les solicitó que con la variable escogida determinaran las medidas de tendencia central tanto para datos agrupados como para datos no agrupados y que realizaran la respectiva interpretación.

El objetivo de la tercera sesión era calcular e interpretar las medidas de dispersión (Varianza, desviación estándar y coeficiente de variación) para datos agrupados y no agrupados de acuerdo a la variable escogida en las sesiones anteriores. Finalmente, el objetivo de la cuarta sesión correspondía al cálculo y la interpretación de las medidas de posición, donde también se realizó la construcción e interpretación de diagramas de cajas y bigotes.

Nota: Las sesiones 3 y 4 fueron realizadas con hoja de cálculo de Excel y de manera manual, con el fin de realizar una comparación de los resultados y mostrarles a los estudiantes las diferentes herramientas que son útiles al momento de realizar cálculos con grandes bases de datos, una vez los estudiantes entendieron su funcionamiento se les dio la opción de continuar trabajando con los computadores.

### **3.2. El proyecto: Las estadísticas de la prueba saber año 2017.**

Este proyecto toma como referencia el trabajo desarrollado por Rouncenfield citado en (Batanero, Díaz, & Gea, Las estadísticas de pobreza y desigualdad, 2011), y los resultados de las reuniones con el director del presente trabajo de grado para la elaboración y estructura de las actividades.

Realizar la lectura, interpretación y análisis de datos estadísticos que se presentan con los puntajes obtenidos en la Prueba Saber año 2017 en Colombia, comparando la relación que existe entre las diferentes variables.

### 3.2.2. Los datos.

La siguiente tabla (ver **Tabla 3.15** Datos de los resultados de la prueba saber año 2017) presenta un bosquejo de los resultados de la prueba saber año 2017, donde se muestra algunas de las 15 variables seleccionadas que se utilizaron para el desarrollo de la actividad (Ver: Anexo: Hojas de códigos. Censo Habitantes de calle Bogotá año 2017, Estadísticas vitales de Colombia año 2017 y Estadísticas de la prueba Saber 11 año 2017)

[illegible]

### 3.2.3. Desarrollo de la actividad

La información que se extrajo para el desarrollo de esta actividad es la correspondiente a la sub-base de datos de los resultados de la prueba saber 11 año 2017, dicha información fue enviada por los correos institucionales de cada uno de los estudiantes junto con la hoja de códigos y las preguntas del proyecto (Anexo: Proyecto. Estadísticas de las pruebas saber año 2017).

Los estudiantes trabajan durante tres sesiones, en grupos de máximo 3 estudiantes, cada grupo tiene asignada tres variables, para luego entregar un informe con la libertad de realizarlo a mano o en Word utilizando herramientas de Excel que se explicaron en las sesiones 3 y 4 previas a la implementación del proyecto.

#### **Análisis a priori de las preguntas del proyecto.**

A continuación, se presenta un análisis a priori para cada una de las preguntas del proyecto teniendo en cuenta el objetivo, el concepto de medida descriptiva a desarrollar y una posible solución. Para este caso utilizaremos las variables que corresponde al grupo A y únicamente tomaremos a V3 como propuesta de solución.

Tabla 3.16 Ítem 1 actividad del proyecto.

***Ítem 1. Determinar para cada una de las variables asignadas la naturaleza (continua, discreta) y su escala (razón, intervalo).***

**Tema: Tipos de variables.**

**Objetivo: Identificar los tipos de variables la naturaleza y su escala.**

**Propuesta de solución:**

**Se escogerá un grupo arbitrariamente, para este caso tomaremos el grupo A; sus variables son: V3 correspondiente a la edad, V8 correspondiente al puntaje de lengua castellana y V10 correspondiente al puntaje de matemáticas, la siguiente tabla resume (ver Tabla 3.17 ) una propuesta de solución que resume la información solicitada.**

**Tabla 3.17 Tipos de variables del grupo A.**

<i>ID</i>	<i>Etiqueta</i>	<i>Naturaleza</i>	<i>Escala</i>
V3	Edad	Cuantitativa continua	Razón
V8	Puntuación LC	Cuantitativa discreta	Intervalo
V10	Puntuación MAT	Cuantitativa discreta	Intervalo

Tabla 3.18 Ítem 2 del proyecto.

**Ítem 2. Escojan dos de las variables y determinen el mínimo, máximo,  $Q_1$ ,  $Q_2$ ,  $Q_3$ ,  $\bar{x}$ ,  $s^2$  y  $s$ .**

**Tema: Media aritmética, medidas de posición (Cuartiles), varianza y desviación estándar.**

**Objetivo: Calcular las medidas establecidas para las dos variables escogidas.**

**Propuesta de solución:**

Para el caso que las variables escogidas sean V3 y V8 se propone la siguiente solución, aunque como se mencionó anteriormente solo se realizará la solución detallada para la variable edad (V3).

V3

Mínimo: 15 años.

Máximo: 71 años.

Para este caso se propone realizar una tabla de distribución de frecuencias como la que se presenta a continuación, para realizar el cálculo de las medidas solicitadas.

**Tabla 3.19 Tabla de distribución de frecuencias de la variable V3**

<b>V3</b> <b>EDAD</b>	$f_i$	$F_i$	$h_i$	$H_i$
15	15	15	0,20	0,20
16	17	32	0,23	0,43
17	20	52	0,27	0,69
18	9	61	0,12	0,81
19	4	65	0,05	0,87

21	1	66	0,01	0,88
22	1	67	0,01	0,89
23	1	68	0,01	0,91
26	1	69	0,01	0,92
32	1	70	0,01	0,93
33	1	71	0,01	0,95
40	1	72	0,01	0,96
54	1	73	0,01	0,97
55	1	74	0,01	0,99
71	1	75	0,01	1,00
<b>TOTALES</b>	<b>75</b>		<b>1,0000</b>	

Para este caso se propone determinar los cuartiles para una tabla de distribución de datos no agrupados. De acuerdo a lo expuesto en la sección 2.2.4

$$Q_1 = P_{25}$$

$$Q_3 = P_{75}$$

Para hallar el percentil correspondiente determinaremos el índice  $i$  que determina la posición, para calcularlo utilizaremos con la fórmula:

$$i = \frac{k n}{100} \quad (2-19)$$

Es decir,

$$i = \frac{25 \times 75}{100} = 18.75 \approx 19$$

Una vez encontrado el índice o la posición del cuartil, se determina el valor del  $Q_1$  observando el valor en la tabla de distribución de frecuencias. Para este caso el valor del  $Q_1$  es 16 años. Con este mismo procedimiento se encuentran los valores de  $Q_2$  y  $Q_3$ , que corresponden a los valores de 17 años y 18 años respectivamente.

En cuanto al valor de  $\bar{x}$ , se puede determinar utilizando la tabla de distribución de frecuencia de la siguiente manera:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i f_i$$

$$\bar{x} = \frac{15 \times 15 + 16 \times 17 + 17 \times 20 + 18 \times 9 + \dots + 24 \times 1 + 32 \times 1 + \dots + 71 \times 1}{75}$$

$$\bar{x} = 19.36$$

Para determinar los valores de  $s^2$  y  $s$ , nuevamente utilizamos la tabla de distribución de frecuencias y el valor encontrado de  $\bar{x}$

$$s^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot f_i}{n - 1}$$

$$s^2 = \frac{(15 - 19.36)^2 \times 15 + (16 - 19.36)^2 \times 17 + \dots + (71 - 19.36)^2 \times 1}{75 - 1}$$

$$s^2 \approx 87.75$$

$$s \approx 9.36$$

Tabla 3.20 Ítem 3 del proyecto.

**Ítem 3. Interpreten  $Q_1$  y  $\bar{x}$ , en las variables escogidas.**

**Tema: Cuartiles y media aritmética.**

**Objetivo:** Interpretar los resultados obtenidos de las medidas de posición y de las medidas de tendencia central en particular el de la media aritmética  $\bar{x}$ .

**Propuesta de solución:**

Con los resultados obtenidos en la variable V3 en el ítem anterior para los valores  $Q_1$  y  $\bar{x}$ , se realiza la interpretación correspondiente.

Para  $Q_1$  en la V3 se obtuvo como resultado 16 años, esto quiere decir que al menos el 25% de la población tiene una edad menor o igual que 16 años, y que al menos el 75% de la población tiene una edad mayor o igual a 16 años.

**El resultado de  $\bar{x}$  fue de 19.36 esto quiere decir que el promedio de edad de las personas es 19.36 años**

Tabla 3.21 Ítem 4 del proyecto

**Ítem 4. Discutan sobre la moda en cada una de las variables y coloquen sus conclusiones.**

**Tema: La moda.**

**Objetivo: Identificar la moda como el valor de mayor frecuencia en una distribución de datos.**

**Propuesta de solución:**

**Análogo a la solución del ítem 2, se debe construir la tabla de distribución de frecuencias de datos no agrupados para la variable 8 y la variable 10.**

Tabla 3.22 Tabla de distribución de frecuencias de la variable V8.

<b>V8</b> <b>Puntaje LC</b>	<b><math>f_i</math></b>	<b><math>F_i</math></b>	<b><math>h_i</math></b>	<b><math>H_i</math></b>
0	1	1	0,0133	0,0133
28	1	2	0,0133	0,0266
30	1	3	0,0133	0,0400
33	1	4	0,0133	0,0533
34	1	5	0,0133	0,0666
35	1	6	0,0133	0,0800
36	3	9	0,0400	0,1200
37	1	10	0,0133	0,1333
38	1	11	0,0133	0,1466
39	8	19	0,1067	0,2533
40	2	21	0,0267	0,2800
41	2	23	0,0267	0,3066
42	1	24	0,0133	0,3200
43	1	25	0,0133	0,3333
44	2	27	0,0267	0,3600
45	2	29	0,0267	0,3866



46	1	30	0,0133	0,4000
47	2	32	0,0267	0,4266
48	2	34	0,0267	0,4533
49	2	36	0,0267	0,4800
51	5	41	0,0667	0,5466
52	1	42	0,0133	0,5600
53	3	45	0,0400	0,6000
54	3	48	0,0400	0,6400
55	1	49	0,0133	0,6533
56	1	50	0,0133	0,6666
57	2	52	0,0267	0,6933
59	2	54	0,0267	0,7200
60	1	55	0,0133	0,7333
61	3	58	0,0400	0,7733
62	3	61	0,0400	0,8133
63	2	63	0,0267	0,8400
64	1	64	0,0133	0,8533
65	2	66	0,0267	0,8800
66	2	68	0,0267	0,9066
69	3	71	0,0400	0,9466
71	2	73	0,0267	0,9733
72	2	75	0,0267	1,0000
TOTALES	75		1,0000	

**Tabla 3.23** Tabla de distribución de frecuencias de la variable V10

<b>V10</b> <b>Puntaje de MAT</b>	<b><math>f_i</math></b>	<b><math>F_i</math></b>	<b><math>h_i</math></b>	<b><math>H_i</math></b>
0	1	1	0,0133	0,0133
26	1	2	0,0133	0,0266
27	1	3	0,0133	0,0400
28	1	4	0,0133	0,0533
29	2	6	0,0267	0,0800
31	1	7	0,0133	0,0933

32	2	9	0,0267	0,1200
33	2	11	0,0267	0,1466
35	2	13	0,0267	0,1733
36	4	17	0,0533	0,2266
37	2	19	0,0267	0,2533
38	4	23	0,0533	0,3066
39	2	25	0,0267	0,3333
40	1	26	0,0133	0,3466
42	2	28	0,0267	0,3733
43	1	29	0,0133	0,3866
44	1	30	0,0133	0,4000
46	3	33	0,0400	0,4400
47	3	36	0,0400	0,4800
48	1	37	0,0133	0,4933
49	2	39	0,0267	0,5200
50	2	41	0,0267	0,5466
51	2	43	0,0267	0,5733
53	2	45	0,0267	0,6000
54	3	48	0,0400	0,6400
56	1	49	0,0133	0,6533
57	3	52	0,0400	0,6933
58	1	53	0,0133	0,7066
59	3	56	0,0400	0,7466
60	2	58	0,0267	0,7733
61	3	61	0,0400	0,8133
62	1	62	0,0133	0,8266
64	3	65	0,0400	0,8666
65	3	68	0,0400	0,9066
66	1	69	0,0133	0,9200
67	1	70	0,0133	0,9333
69	2	72	0,0267	0,9600
71	1	73	0,0133	0,9733
72	1	74	0,0133	0,9866
76	1	75	0,0133	1,0000

<b>TOTALES</b>	<b>75</b>	<b>1,0000</b>
----------------	-----------	---------------

Haciendo la revisión de cada tabla de distribución es posible establecer que la moda para la variable V3 (Ver Tabla 3.19) es igual a 17 años ya que su valor tiene una frecuencia absoluta ( $f_i$ ) de 20 datos. En cuanto a la variable V8 (Ver Tabla 3.22) la moda es el puntaje de 39, con una frecuencia absoluta ( $f_i$ ) de 8 datos. Finalmente la variable V10 (ver Tabla 3.23) es una variable con moda bimodal ya que presenta puntajes de 36 y 38 ambos con frecuencia absoluta ( $f_i$ ) de 4 datos.

Tabla 3.24 Ítem 5 del proyecto.

**Ítem 5. Realiza el boxplot con alguna de las variables ¿Hay datos atípicos?****Tema: Boxplot (Diagramas de cajas y bigotes)**

**Objetivo:** Construir un boxplot para identificar posibles valores atípicos con respecto a una de las variables.

**Propuesta de solución:**

Una manera de determinar los valores atípicos de la variable V10, es utilizar un diagrama de cajas y bigotes, para la construcción del gráfico se debe tener en cuenta lo descrito en las sesión 2.2.6.2 gráficos de variables cuantitativas.

- Cuartiles.

Primer cuartil ( $Q_1$ ) para este se tiene en cuenta lo descrito en la sesión 2.2.4.

Donde se establece que  $Q_1 = P_{25}$ . Por lo tanto, el índice o posición del cuartil es igual a.

$$i = \frac{k \cdot n}{100} = \frac{25 \times 75}{100} = 18.75 \approx 19$$

Una vez encontrada la posición del cuartil, se determina el valor del  $Q_1$ , observando el valor de la variable en la tabla de distribución de frecuencias. Que en este caso es 37 puntos. De manera análoga se calcula el segundo y tercer cuartil,  $Q_2 = 49$  y  $Q_3 = 60$ .

- Rango intercuartílico corresponde a  $RI_x = Q_3 - Q_1 = 23$

- El límite inferior del gráfico corresponde a  $LI = Q_1 - 1.5 \cdot RI_x = 2.5$  y el bigote inferior de la caja  $bigote.inf = \max\{x_{min}, Q_1 - 1.5 \cdot RI_x\} = 26$
- El límite superior del gráfico corresponde a  $LS = Q_3 + 1.5 \cdot RI_x = 94.5$  y el bigote superior de la caja  $bigote.sup = \min\{x_{max}, Q_3 + 1.5 \cdot RI_x\} = 76$

Finalmente, se elabora el boxplot para el puntaje de matemáticas (V10)

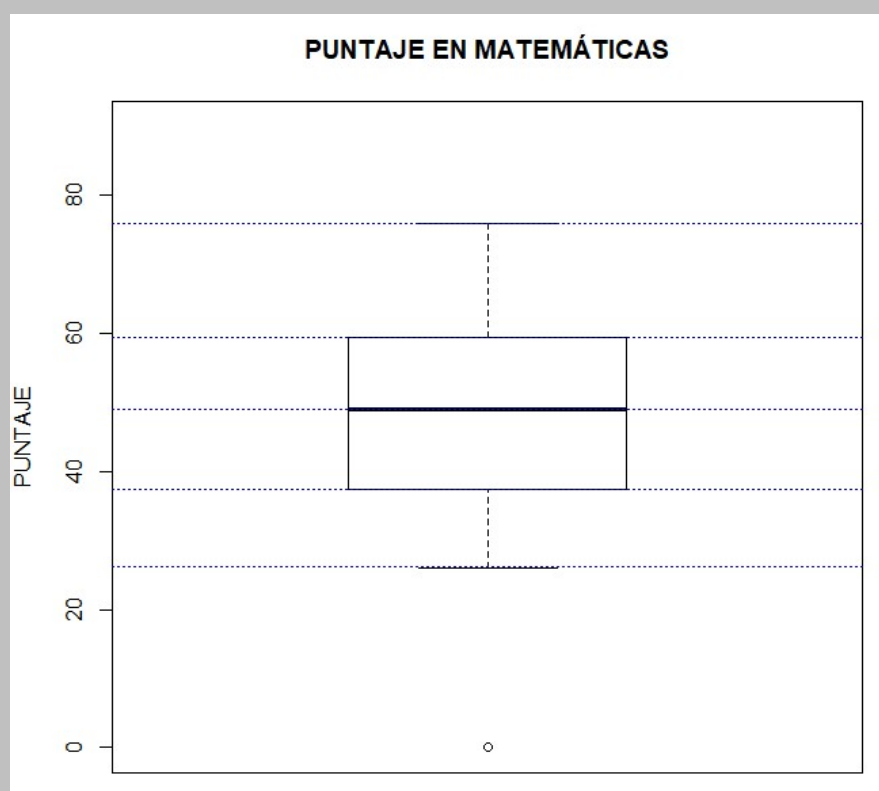


Figura 3-1 Boxplot de la variable puntaje de matemáticas

Tabla 3.25 Ítem 6 del proyecto.

**Ítem 6. Realizar el histograma y el polígono de frecuencias de la variable V3.**

**Tema:** Histograma y polígono de frecuencias.

**Objetivo:** Realizar un gráfico estadístico (histograma) para representar las edades V3

**Propuesta de solución:**

Para la construcción del histograma, se realizan los siguientes procedimientos.

- El rango de edades es  $R_X = x_{max} - x_{min} = 71 - 15 = 56$ .
- Para este caso se escogen 5 intervalos ya que con la regla de Sturges,  $k = \lceil 1 + 3.32 \log_{10} n \rceil = \lceil 1 + 3.32 \log_{10} 75 \rceil$  aparecen 8 intervalos si se considera este caso quedan intervalos con  $f_i$  igual a 0
- La amplitud de cada intervalo es  $A = \frac{R_X}{k} = \frac{56}{5} = 11.2$ .

La siguiente tabla (ver Tabla 3.26 ) muestra la distribución de frecuencias de datos agrupados para V3.

Tabla 3.26 Distribución de frecuencias de datos agrupados para V3

$k_i$	$[x_{i-1}, x_i)$	$X'_i$	$f_i$	$F_i$	$h_i$	$H_i$
1	[15    26.2	20.6	69	69	0.92	0.92
2	26.2    37.4	31.8	2	71	0.03	0.95
3	37.4    48.6	43	1	72	0.01	0.96
4	48.6    59.8	54.2	2	74	0.03	0.99
5	59.8    71]	65.4	1	75	0.01	1.00
<b>TOTALES</b>			75		1,00	

Finalmente, se elabora el histograma y el polígono de frecuencias que representan la variable edad (V3).

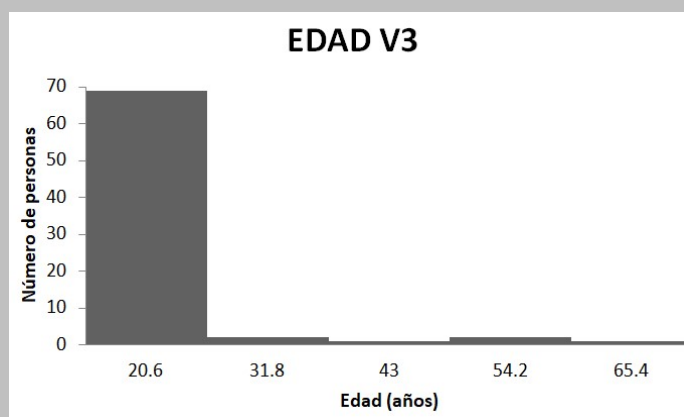
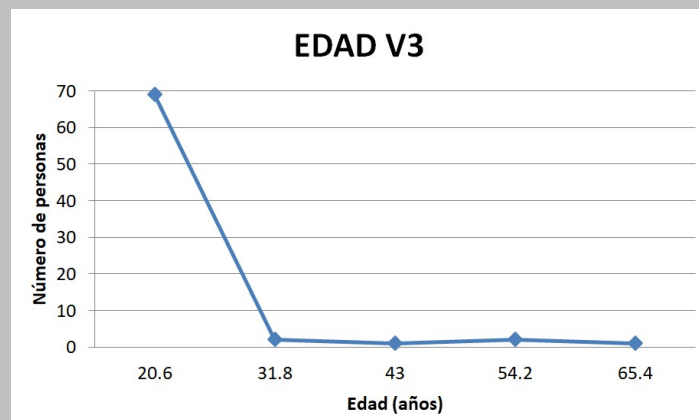


Figura 3-2 Histograma de la variable edad V3.



**Figura 3-3 Polígono de frecuencias de la variable edad V3.**

## 4. Estrategia didáctica: Resultados y análisis

En este capítulo se analiza la información que se obtuvo, con respecto a los resultados o producciones de los cuatro grupos de estudiantes que se mencionaron en la sección 1.1, en cuanto a la prueba diagnóstica, como en el desarrollo del proyecto estadísticas de la pruebas saber año 2017 y la prueba final.

### 4.1. Prueba diagnóstica: Análisis de resultados

La prueba diagnóstica inicialmente se aplica a 98 de los 120 estudiantes de grado décimo, debido a que por diferentes situaciones particulares algunos estudiantes no les fue posible presentar la prueba. El tiempo de ejecución de la prueba fue un bloque de clase, que corresponde a 90 minutos. El objetivo de la prueba era identificar los conocimientos previos. Luego, como un segundo momento, después de la implementación del proyecto: “Estadísticas de las pruebas saber del año 2017”, aplicar la prueba nuevamente con algunas modificaciones que se realizaron en algunas preguntas. Esta prueba, denominada “prueba final” (Ver Anexo: Prueba final.), se aplicó durante 90 minutos para poder determinar los conocimientos adquiridos por los estudiantes evaluados con la “prueba diagnóstica”.

Es importante aclarar que los ajustes que se realizaron a la prueba inicial fueron principalmente por dos razones, la primera, porque de acuerdo a lo manifestado por la mayoría de los estudiantes en los cuatro grupos, para ellos no fue clara la pregunta número 2 (ver **Tabla 3.3**); lo que dificulta el análisis de la pregunta. La segunda razón tiene que ver con el tiempo de duración de la prueba, ya que para la pregunta número 10 (ver **Tabla 3.11** y **Tabla 3.12**) la cantidad de datos inicialmente colocados hizo que los estudiantes que realizaron los cálculos tardaran mucho más tiempo; aunque en particular esta pregunta no fue resuelta por la mayoría de los estudiantes, fue evidente la falta de tiempo para intentar

resolverla. Más adelante se muestra de manera más detallada lo sucedido con estas dos preguntas.

#### **4.1.1. Análisis cualitativo, prueba diagnóstica, previo a la implementación del proyecto.**

Inicialmente para el desarrollo de este proyecto, se aplicó la prueba diagnóstica, que se describe detalladamente en la sección 3.1.1. Los resultados de esta prueba permitieron establecer aspectos a tener en cuenta previos al desarrollo del proyecto (Estadísticas de la prueba Saber año 2017), tales como el diseño de las actividades de las sesiones, (ver sección 3.1.2) cuya finalidad es afianzar los conceptos de medidas descriptivas y resolver las dudas que se evidenciaron en el desarrollo de la prueba diagnóstica.

Durante el desarrollo de la prueba se observaron dificultades con respecto a: 1) La interpretación de diagramas para obtener información y poder calcular las medidas de tendencia central; 2) No comprender el concepto de frecuencia relativa; 3) No reconocer en los intervalos, donde inicia y donde finaliza. Es decir, dificultades con la interpretación simbólica de intervalos abiertos y cerrados. 4) No recordar cómo se trabaja con porcentajes y cómo establecer las cantidades que representan; 5) No interpretar gráficos como el diagrama de cajas y bigotes. También es importante mencionar que los estudiantes manifestaron preguntas durante el desarrollo de la prueba como, por ejemplo

¿Cómo así la frecuencia relativa? ¿Ese gráfico (boxplot) qué quiere decir? ¿Cuál es la media aritmética? ¿Cuáles son los cuartiles? ¿Cómo se calcula la varianza?

Teniendo como referente estas observaciones se realiza una revisión de los resultados de manera general en los cuatro grupos. La Tabla 4.1, Tabla 4.2, Tabla 4.3 y Tabla 4.4 presenta la razón de aciertos en cada una de las preguntas y el porcentaje de aciertos para cada uno de los grupos.



Tabla 4.1 Resultados de la prueba diagnóstica grupo 1001

<b>Pregunta</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10ª</b>	<b>10b</b>
<i>Cantidad de aciertos en el grupo</i>	$\frac{22}{24}$	$\frac{8}{24}$	$\frac{22}{24}$	$\frac{24}{24}$	$\frac{24}{24}$	$\frac{16}{24}$	$\frac{22}{24}$	$\frac{21}{24}$	$\frac{5}{24}$	$\frac{17}{24}$	$\frac{0}{24}$
<i>Porcentaje de aciertos</i>	92%	33%	92%	100%	100%	67%	92%	88%	21%	71%	0%

En el grupo 1001 se observó que las preguntas que generaron mayor dificultad son: **2, 6, 9, 10b.**

Tabla 4.2 Resultados de la prueba diagnóstica grupo 1002

<b>Pregunta</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10ª</b>	<b>10b</b>
<i>Cantidad de aciertos en el grupo</i>	$\frac{24}{26}$	$\frac{17}{26}$	$\frac{25}{26}$	$\frac{25}{26}$	$\frac{25}{26}$	$\frac{21}{26}$	$\frac{24}{26}$	$\frac{23}{26}$	$\frac{7}{26}$	$\frac{14}{26}$	$\frac{2}{26}$
<i>Porcentaje de aciertos</i>	92%	65%	96%	96%	96%	81%	92%	88%	27%	54%	8%

En el grupo 1002 se observó que las preguntas que generaron mayor dificultad son: **2, 9, 10a, 10b.**

Tabla 4.3 Resultados de la prueba diagnóstica grupo 1003

<b>Pregunta</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10ª</b>	<b>10b</b>
<i>Cantidad de aciertos en el grupo</i>	$\frac{24}{25}$	$\frac{14}{25}$	$\frac{24}{25}$	$\frac{23}{25}$	$\frac{25}{25}$	$\frac{11}{25}$	$\frac{22}{25}$	$\frac{22}{25}$	$\frac{10}{24}$	$\frac{15}{25}$	$\frac{1}{24}$
<i>Porcentaje de aciertos</i>	96%	56%	96%	92%	100%	44%	88%	88%	40%	60%	4%

En el grupo 1003 se observó que las preguntas que generaron mayor dificultad son: **2, 6, 9, 10a, 10b.**

Tabla 4.4 Resultados de la prueba diagnóstica grupo 1004

<b>Pregunta</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10ª</b>	<b>10b</b>
<i>Cantidad de aciertos en el grupo</i>	$\frac{24}{24}$	$\frac{10}{24}$	$\frac{22}{24}$	$\frac{22}{24}$	$\frac{23}{24}$	$\frac{23}{24}$	$\frac{23}{24}$	$\frac{24}{24}$	$\frac{8}{24}$	$\frac{17}{24}$	$\frac{2}{24}$
<i>Porcentaje de aciertos</i>	100%	42%	92%	92%	96%	96%	96%	100%	33%	71%	8%

En el grupo 1004 se observó que las preguntas que generaron mayor dificultad son: **2, 9, 10b.**

De manera general se evidencia que las preguntas en donde se reconoce mayor dificultad son las preguntas 2, 6, 9, 10a, 10b, así que con el fin de establecer cuáles fueron las posibles dificultades que pudieron incidir en los resultados desfavorables de dichas preguntas se plantea la **Tabla 4.5**, que muestra las posibles causas de las dificultades.

**Tabla 4.5** Análisis cualitativo de las dificultades en el desarrollo de la prueba diagnóstica.

<b>Pregunta</b>	<b>Dificultades</b>
2	Los estudiantes no tienen claro el concepto de frecuencia y tampoco lo relacionan cuando está representado en un gráfico, esta pregunta requiere de este concepto ya que luego esta información la utilizan para calcular el promedio. También se observa que el gráfico que se presentó no es el apropiado porque puede ser confundido con un histograma.
6	De acuerdo a lo observado y lo manifestado por algunos de ellos, no se reconocen expresiones como “ <i>edad mayor o igual que 21 años y menor de 25 años</i> ” y no encuentran como relacionarlo con los límites de los intervalos que aparecen en la tabla. Por otra parte, también se observa que no establecen la relación entre la frecuencia relativa y los porcentajes, lo que no permitía que hicieran los cálculos respectivos.
9	Se evidencia que muchos de los estudiantes no reconocen un gráfico de cajas y bigotes (Boxplot), algunos manifestando que nunca lo habían visto y otros diciendo que no recordaban bien cómo funciona y que información se puede deducir de este. Cuando se indagó a los estudiantes que contestaron acertadamente a las preguntas manifestaron que lo habían realizado al azar que tampoco conocían mucha información al respecto.
10 a	Aunque esta pregunta de acuerdo a los resultados no es la que presenta mayor dificultad, si se observó que algunos estudiantes no tienen claro el concepto de media aritmética (promedio) y mediana, como también el confundir cuál es la que representa cada una de ellas y como hacer los cálculos correspondientes en cada una.
10 b	Es la pregunta donde se presenta mayor dificultad en los cuatro grupos debido a que casi todos los estudiantes manifestaron que no sabían o no recordaban que era la varianza y las medidas de posición, ni como

#### 4.1.2. Análisis cualitativo, prueba final, después de implementar el proyecto.

Tabla 4.6 Resultados de la prueba final grupo 1001.

**Tabla 4.7** Resultados de la prueba final grupo 1002.

[illegible]

**Tabla 4.8** Resultados de la prueba final grupo 1003.

<b>Pregunta</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10a</b>	<b>10b</b>
<i>Cantidad de aciertos en el grupo</i>	$\frac{23}{26}$	$\frac{19}{26}$	$\frac{26}{26}$	$\frac{22}{26}$	$\frac{25}{26}$	$\frac{25}{26}$	$\frac{26}{26}$	$\frac{26}{26}$	$\frac{21}{26}$	$\frac{23}{26}$	$\frac{22}{26}$
<i>Porcentaje de aciertos</i>	88%	73%	100%	85%	96%	96%	100%	100%	81%	88%	85%

**Tabla 4.9** Resultados de la prueba final grupo 1004.

<b>Pregunta</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10a</b>	<b>10b</b>
<i>Cantidad de aciertos en el grupo</i>	$\frac{26}{29}$	$\frac{24}{29}$	$\frac{29}{29}$	$\frac{27}{29}$	$\frac{28}{29}$	$\frac{28}{29}$	$\frac{29}{29}$	$\frac{29}{29}$	$\frac{26}{29}$	$\frac{28}{29}$	$\frac{23}{29}$
<i>Porcentaje de aciertos</i>	90%	83%	100%	93%	97%	97%	100%	100%	90%	97%	79%

Haciendo una comparación de los resultados a nivel de grupo, se observa que obtuvieron mejores resultados de la pregunta 5 en adelante. Las preguntas con mejores resultados son las preguntas **9**, **10a** y **10b**. Por otra parte, hubo preguntas como la **1** y **2** donde el desempeño fue menor que el anterior, para el caso de la pregunta 1 el desempeño bajo un poco con respecto a la prueba inicial en los 4 grupos, en el caso de la pregunta 2, en tres de los cuatro grupos los resultados desmejoraron con respecto a la prueba inicial. Igualmente hay unos casos especiales en las preguntas **3** y **4** del curso 1001, en donde estas dos preguntas desmejoraron con respecto a los resultados de los otros tres grupos.

Estos resultados hacen pensar que la estrategia didáctica que se implementó ha generado un efecto favorable en los estudiantes sobre todo en las preguntas **9**, **10a** y **10b**, donde la mejora fue significativa. A continuación, se presenta un análisis de resultados para los cuatro grupos donde se hace una comparativa pregunta por pregunta, para establecer la apreciación con respecto a la pertinencia de la estrategia didáctica (Proyecto Estadísticas de la prueba Saber 2017).

### 4.1.3. Análisis estadístico de la prueba diagnóstica, antes y después de aplicar el proyecto Estadísticas de la prueba Saber 2017.

El siguiente análisis estadístico se desarrolló con 90 de 120 estudiantes, esto se dio porque solo con este número de estudiantes se pudo obtener los resultados de las dos pruebas. Para el análisis de resultados presentaremos una tabla de contingencia de  $2 \times 2$  (ver Tabla 4.10) en donde colocaremos los resultados del antes y después de las pruebas, las preguntas correctas las codificaremos con uno y las incorrectas con un cero.

Tabla 4.10. Presentación de los datos.

		Después		Total	Probabilidades	
Antes	Pregunta	1	0			
			1			
	0	a	b	a + b	Ganancia	$\frac{b}{a+b}$
	1	c	d	c + d	En contra	$\frac{c}{c+d}$
Total						

En la Tabla 4.10 también se definen las siguientes probabilidades de interés:

$$P_{ganancia} = P_g = \frac{b}{a+b}$$

$$P_{en\ contra} = P_c = \frac{c}{c+d}$$

Donde  $P_g$  representa la probabilidad de pasar de respuesta incorrecta, antes de la estrategia didáctica, a respuesta correcta, después de la estrategia didáctica, en la misma pregunta. Para el caso de la probabilidad en contra  $P_c$  representa la probabilidad de pasar de respuesta correcta, antes de la estrategia didáctica, a respuesta incorrecta, para la misma pregunta.

#### 4.1.3.1. Conclusiones por pregunta.

Tabla 4.11. Datos de la prueba. Pregunta 1.

Pregunta	1	Después		Total	Probabilidades	
		0	1			
Antes	0	0	5	5	Ganancia	$\frac{5}{5} = 1$
	1	9	76	85	En contra	$\frac{9}{85} = 0.106$
	Total			90		

**Conclusión:** La probabilidad de ganancia  $P_g = 100\%$ , esto quiere decir que los cinco estudiantes que marcaron respuesta incorrecta lograron pasar a respuesta correcta. La probabilidad en contra  $P_c = 10.6\%$  muestra que 9 estudiantes desmejoraron. Es decir, pasaron de estar correctos a incorrectos.

Tabla 4.12. Datos de la prueba. Pregunta 2.

Pregunta	2	Después		Total	Probabilidades	
		0	1			
Antes	0	13	34	47	Ganancia	$\frac{34}{47} = 0.723$
	1	6	37	43	En contra	$\frac{6}{43} = 0.139$
	Total			90		

**Conclusión:** La probabilidad de ganancia  $P_g = 72.3\%$ . Es decir, de 47 estudiantes que marcaron respuesta incorrecta 34 lograron pasar a respuesta correcta y 13 estudiantes siguen marcando respuesta incorrecta. La probabilidad en contra  $P_c = 13.9\%$  muestra que seis estudiantes desmejoraron. Por tal motivo se puede decir que hubo ganancia, pero no tan significativa como en la pregunta anterior. Además, los estudiantes manifiestan que la información de la pregunta no era clara y que no pudieron establecer como determinar la solución. Es posible que se presente esta situación porque la pregunta se cambió completamente con respecto a la que se aplicó en la prueba inicial, además el gráfico no fue el más adecuado para presentar la información.

Tabla 4.13. Datos de la prueba. Pregunta 3.

Pregunta	3	Después		Total	Probabilidades	
		0	1			
Antes	0	0	6	6	Ganancia	$\frac{6}{6} = 1$
	1	1	83	84	En contra	$\frac{1}{84} = 0.011$
	Total			90		

**Conclusión:** La probabilidad de ganancia,  $P_g = 100\%$ . Es decir, 6 estudiantes que marcaron respuesta incorrecta lograron pasar a respuesta correcta. La probabilidad en contra  $P_c = 1,1\%$  esto muestra que 1 estudiante desmejoró, pero esto no es una cifra considerable. Es decir que la estrategia didáctica fue efectiva.

Tabla 4.14. Datos de la prueba. Pregunta 4.

Pregunta	4	Después		Total	Probabilidades	
		0	1			
Antes	0	1	4	5	Ganancia	$\frac{4}{5} = 0.8$
	1	5	80	85	En contra	$\frac{5}{80} = 0.06$
	Total			90		

**Conclusión:** La probabilidad de ganancia  $P_g = 80\%$ . Es decir, que de 5 estudiantes que tenían respuesta incorrecta 4 pasaron a respuesta correcta. En el caso de la probabilidad en contra  $P_c = 6\%$  únicamente un estudiante marcó incorrecto y otro sigue estando incorrecto. Es decir, no se puede establecer una desmejora significativa de la estrategia didáctica.

Tabla 4.15. Datos de la prueba. Pregunta 5.

Pregunta	5	Después		Total	Probabilidades	
		0	1			
Antes	0	0	3	3	Ganancia	$\frac{3}{3} = 1$
	1	2	85	87	En contra	$\frac{2}{87} = 0.02$
	Total			90		

**Conclusión:** : La probabilidad de ganancia  $P_g = 100\%$ . Los 3 estudiantes que inicialmente estaban con respuesta incorrecta pasaron a correcta. En el caso de la probabilidad en contra  $P_c = 2\%$  únicamente dos estudiantes marcaron incorrecto. Es decir, no se puede establecer una desmejora significativa de la estrategia didáctica.

Tabla 4.16. Datos de la prueba. Pregunta 6.

Pregunta	6	Después		Total	Probabilidades	
		0	1			
Antes	0	0	25	25	Ganancia	$\frac{25}{25} = 1$
	1	2	63	65	En contra	$\frac{2}{63} = 0.03$
	Total			90		

**Conclusión:** La probabilidad de ganancia,  $P_g = 100\%$  Los 25 estudiantes que inicialmente estaban con respuesta incorrecta pasaron a correcta, es decir que la estrategia didáctica fue efectiva. Para el caso de la probabilidad en contra,  $P_c = 3\%$  únicamente 2 estudiantes pasaron de estar correctos a incorrectos. Para este caso, por ser una cantidad tan pequeña no se puede establecer si hubo pérdida significativa con la implementación de la estrategia didáctica.

Tabla 4.17. Datos de la prueba 1001 pregunta 7.

Pregunta	7	Después		Total	Probabilidades	
		0	1			
Antes	0	0	9	9	Ganancia	$\frac{1}{1} = 1$
	1	1	80	81	En contra	$\frac{1}{81} = 0.01$
	Total			90		

**Conclusión:** La probabilidad de ganancia,  $P_g = 100\%$ . Los 9 estudiantes que inicialmente estaban con respuesta incorrecta pasaron a correcta, para el caso de la probabilidad en contra,  $P_c = 1\%$  desmejoró un estudiante. Es decir, no se puede establecer una desmejora significativa de la estrategia didáctica.



Tabla 4.18. Datos de la prueba. Pregunta 8.

Pregunta	8	Después		Total	Probabilidades	
		0	1			
Antes	0	0	8	8	Ganancia	$\frac{8}{8} = 1$
	1	0	82	82	En contra	$\frac{0}{82} = 0$
	Total			90		

**Conclusión:** La probabilidad de ganancia,  $P_g = 100\%$ . Los 8 estudiantes que inicialmente estaban con respuesta incorrecta pasaron a correcta, para el caso de la probabilidad en contra,  $P_c = 0\%$  ningún estudiante desmejoró y ahora todos los estudiantes están correctos lo que da significancia a la estrategia didáctica aplicada.

Tabla 4.19. Datos de la prueba. Pregunta 9.

Pregunta	9	Después		Total	Probabilidades	
		0	1			
Antes	0	5	60	65	Ganancia	$\frac{60}{65} = 0.923$
	1	2	23	25	En contra	$\frac{2}{25} = 0.08$
	Total			24		

**Conclusión:** La probabilidad de ganancia,  $P_g = 92\%$  de los 65 estudiantes que inicialmente estaban con respuesta incorrecta 60 pasaron a respuesta correcta, para el caso de la probabilidad en contra,  $P_c = 8\%$  2 estudiantes desmejoraron y 5 aún tienen la respuesta incorrecta. Es decir, no se puede establecer una desmejora significativa de la estrategia didáctica.

Tabla 4.20. Datos de la prueba. Pregunta 10a.

Pregunta	10a	Después		Total	Probabilidades	
		0	1			
Antes	0	5	26	31	Ganancia	$\frac{26}{31} = 0.838$
	1	6	53	59	En contra	$\frac{6}{59} = 0.101$
	Total			90		

**Conclusión:** La probabilidad de ganancia,  $P_g = 83.8\%$ . de los 31 estudiantes que inicialmente estaban con respuesta incorrecta 26 pasaron a respuesta correcta. Para el caso de la probabilidad en contra,  $P_c = 10.1\%$  6 estudiantes desmejoraron pasaron de correcto a incorrecto y 6 aún tienen la respuesta incorrecta.

Tabla 4.21. Datos de la prueba. Pregunta 10b.

Pregunta	10b	Después		Total	Probabilidades	
		0	1			
Antes	0	21	65	86	Ganancia	$\frac{65}{86} = 0.809$
	1	1	3	4	En contra	$\frac{1}{4} = 0.25$
	Total			90		

**Conclusión:** La probabilidad de ganancia es  $P_g = 80.9\%$ , de los 86 estudiantes que inicialmente estaban con respuesta incorrecta, 65 pasaron a correcta y 21 se mantienen con la respuesta incorrecta, para el caso de la probabilidad en contra,  $P_c = 25\%$ , la más alta en todas la 11 preguntas, aunque se puede concluir que es una ganancia significativa porque más de la mitad de los estudiantes mejoraron y solo uno desmejoró.

Para mostrar los resultados de manera general, se realiza un gráfico (ver Figura 4-1) con las preguntas desarrolladas. El gráfico representa las probabilidades de ganancia (Bueno) y en contra (malo), donde es notoria la favorabilidad que tuvo la aplicación del proyecto, debido a que las barras que se encuentran sobre bueno corresponde a la ganancia y la mayoría están 1, como es el caso de las preguntas 1,3, 5, 6, 7 y 8 o están muy próximas a 1.

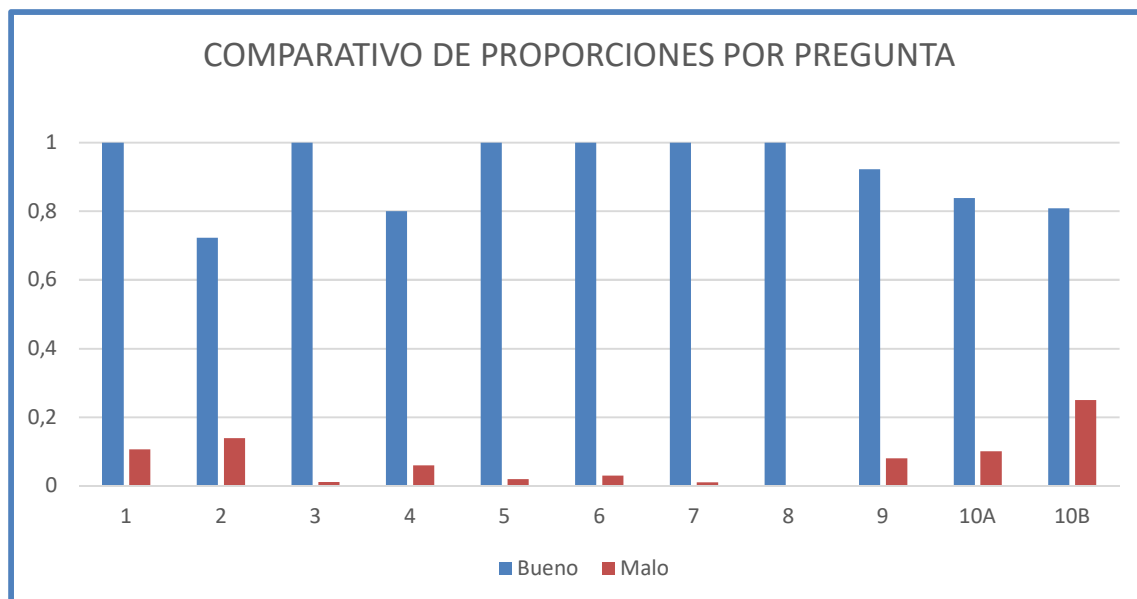


Figura 4-1 Resultados generales en las preguntas una vez aplicada la “prueba final”

En síntesis, podemos decir que el trabajo realizado por la metodología del trabajo por proyectos, deja muchos aspectos positivos debido a que esta estrategia didáctica dio cuenta de los aprendizajes que se generaron en los estudiantes, que a su vez en sus comentarios manifestaron su satisfacción con el trabajo realizado en la estrategia didáctica. Ya que, según ellos, muchos de los contenidos ya eran de su conocimiento, pero con la experiencia quedaron mucho más claros. Por otro lado, fue muy satisfactorio reconocer en la experiencia los conceptos en donde los estudiantes presentaron dificultad, ya que toda la estrategia se diseñó para poder superar las dificultades encontradas y generar diferentes ambientes de aprendizaje que contribuyeran en la formación de los estudiantes.

De manera más puntual, en la estrategia didáctica, en la parte conceptual y procedimental se observa que los estudiantes presentan dificultades al momento de clasificar las variables por su naturaleza, sobre todo cuando las variables son de tipo cuantitativo. De igual manera se evidencia la dificultad para interpretar resultados específicos que se obtienen de las medidas de posición, ya que muchos asocian los índices o posición de los percentiles con el valor que da el cuartil, esto genera una dificultad ya que se toman valores erróneos que no corresponden al valor que tomó la variable en esa posición.

En lo que respecta a las tablas de distribución de frecuencias de datos agrupados, algunos estudiantes construyeron tablas donde algunos intervalos tenían frecuencia absoluta igual a cero, este motivo también a que la construcción de los gráficos no estuviera correcta.

Es importante mencionar que la principal limitante en la ejecución de este proyecto fue el tiempo ya que no fueron suficientes las sesiones para poder generar más espacios de retroalimentación que permitieran aclarar las dudas de los estudiantes que después de haber realizado el proyecto continuaron presentando algún tipo de error.

## **4.2. Proyecto Estadísticas de las pruebas Saber año 2017. Análisis de resultados**

Con la implementación del proyecto *Estadísticas de las pruebas Saber año 2017*, se evalúan los aprendizajes obtenidos por los estudiantes, durante el desarrollo de las clases de estadística descriptiva, y las habilidades de aplicarlas en contexto.

Para el desarrollo del proyecto se realizaron tres sesiones de clase con cada uno de los grupos, cada clase tiene una duración de 90 minutos, en la primera sesión se realiza la conformación de los grupos de máximo tres estudiantes por grupo, se asignan las variables de las bases de datos y la actividad a realizar (Ver 3.2.3) los estudiantes realizaron los cálculos manuales y algunos también lo hicieron con hojas de cálculo de Excel.

Una vez terminada las tres sesiones con todos los cursos, cada grupo de estudiantes debe presentar un informe escrito teniendo en cuenta los resultados y el análisis del proyecto que se desarrolló. A continuación, se describen algunos errores encontrados en cada una de las preguntas.

**Ítem 1:** Determinar para cada una de las variables asignadas la naturaleza (Continua, discreta) y su escala. (Razón, intervalo)

En las respuestas dadas por los estudiantes para el desarrollo del primer ítem del proyecto se encuentran algunos errores como no determinar cuál es la escala de la variable V3 (edad) como se muestra en la Figura 4-2 y Figura 4-3

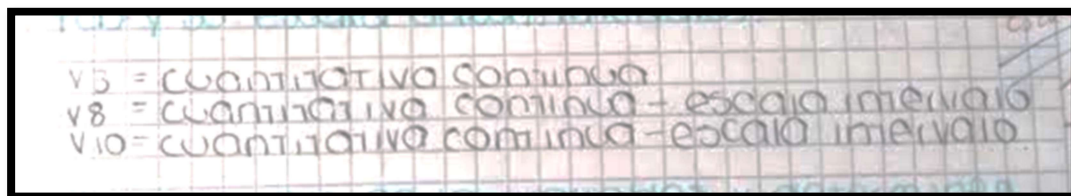


Figura 4-2. Ejemplo de respuesta ítem 1

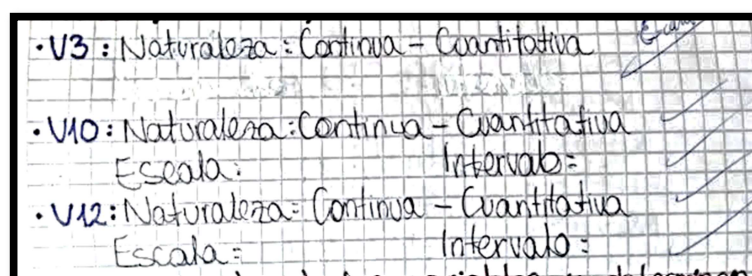


Figura 4-3. Ejemplo 2 respuesta de ítem 1.

Esto muestra una dificultad por parte de algunos estudiantes para reconocer las escalas de una variable en particular, la mayor dificultad manifestada por los estudiantes es la de identificar la escala de una variable de tipo cuantitativo.

**Ítem 2.** Escojan dos de las variables y determinen el mínimo, máximo,  $Q_1$ ,  $Q_2$ ,  $Q_3$ ,  $\bar{x}$ ,  $s^2$  y  $s$ .

Con los resultados obtenidos en los cuatro cursos se pudo identificar que el error que más se cometió fue al momento de calcular los cuartiles, debido a que algunos estudiantes calculaban la posición del cuartil y asumían que este resultado correspondía al valor del cuartil, más no identificaban cual era el valor de la variable que ocupaba esta posición. (Ver Figura 4-4 )

$$\begin{array}{l}
 V12 = Q_1 = \frac{1 \cdot 75}{4} = 18.75 \\
 Q_2 = \frac{2 \cdot 75}{4} = 37.5 \\
 Q_3 = \frac{3 \cdot 75}{4} = 56.25
 \end{array}
 \quad
 \begin{array}{l}
 V3 = Q_1 = \frac{1 \cdot 75}{4} = 18.75 \\
 Q_2 = \frac{2 \cdot 75}{4} = 37.5 \\
 Q_3 = \frac{3 \cdot 75}{4} = 56.25
 \end{array}$$

Figura 4-4. Ejemplo de respuesta ítem 2.

La Figura 4-4 muestra un ejemplo de lo mencionado anteriormente donde únicamente los estudiantes determinaron las posiciones y no el valor de la variable que ocupa esa posición

**Ítem 3.** Interpreten  $Q_1$ ,  $\bar{x}$  en cada una de las variables.

En las respuestas dadas por los estudiantes se evidenció una buena interpretación del cuartil 1 y de la media aritmética, la única dificultad es que la mayoría no realizaron la interpretación de las tres variables asignadas. A continuación, mostramos en la Figura 4-5 un ejemplo de interpretación del cuartil 1  $Q_1$  y la media aritmética  $\bar{x}$ .

★ Interpretaciones:

$Q_1$ : El 25% de la muestra obtuvo un puntaje máximo de 37 en matemáticas.

$\bar{x}$ : El promedio que se obtuvo en la puntuación en matemáticas fue de 48,373.

Figura 4-5. Ejemplo de respuesta ítem 3.

**Ítem 4.** Discutan sobre la moda en cada una de las variables y coloquen sus conclusiones.

De acuerdo a los resultados obtenidos los estudiantes mencionaron cuales eran las modas correspondientes a cada una de las variables, además discutieron si había más de una moda o no, en cada una de las variables, el siguiente es un ejemplo de lo realizado por los estudiantes. (Ver Figura 4-6 )

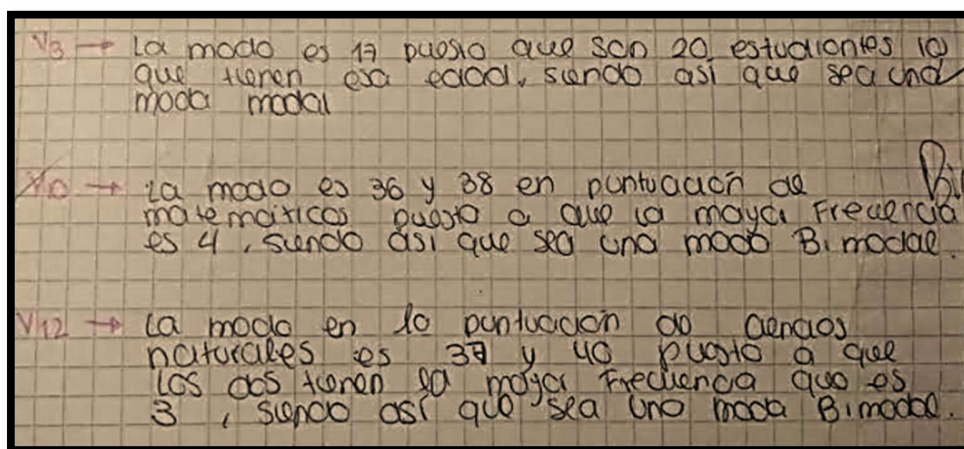


Figura 4-6. Ejemplo de respuesta ítem 4.

**Ítem 5.** Realizar el boxplot con alguna de las variables (No escoger V3) ¿Hay datos atípicos?

Para determinar la existencia de datos atípicos, se les solicitó a los estudiantes la elaboración de un diagrama de cajas y bigotes, con alguna de las variables asignadas menos la variable 3, ya que esta se utilizó como ejemplo en una de las intervenciones. Para el caso del gráfico los estudiantes olvidaron tener en cuenta elementos característicos importantes como la etiqueta del eje  $x$  y el título correspondiente. Algunos presentan errores al seleccionar las escalas y sus gráficos no son claros (Ve Figura 4-7)

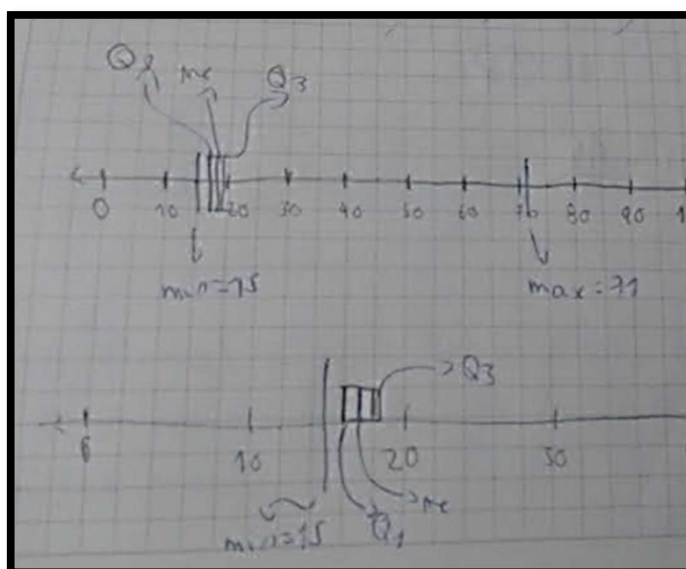


Figura 4-7 Ejemplo de boxplot ítem 5

En otro gráfico realizado por los estudiantes no es posible diferenciar cuales son los datos atípicos y los límites del gráfico, este error se puede dar, por los cálculos que realizaron en los límites (Ver Figura 4-8 Ejemplo 2 boxplot ítem 5. Figura 4-8 )

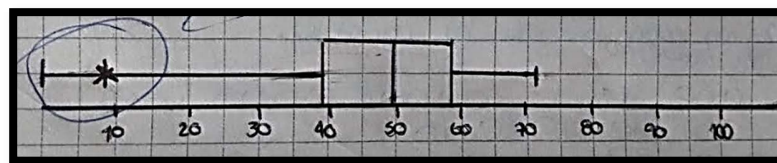


Figura 4-8 Ejemplo 2 boxplot ítem 5.

A continuación, se muestra una representación gráfica que está bien construida pero a la que le hizo falta colocar etiqueta en  $x$  y el título.

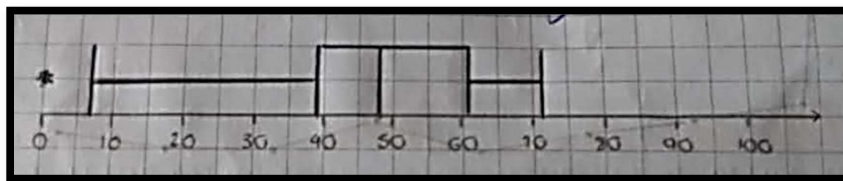


Figura 4-9. Ejemplo 3 boxplot ítem 5.

**Ítem 6.** Realizar el histograma y el polígono de frecuencias de la variable V3.

Para realizar la representación de los datos los estudiantes tenían la información correspondiente a una tabla de distribución de frecuencias de datos agrupados, que anteriormente habían elaborado con dicha información realizaron la correspondiente construcción del histograma y el polígono de frecuencias. Al igual que en el ítem anterior los estudiantes no colocaron etiquetas en los ejes y no colocaron el correspondiente título. Como también se evidencia un error de tipo conceptual, ya que hay intervalos que se consideraron con frecuencia absoluta igual a cero ( Ver Figura 4-10)



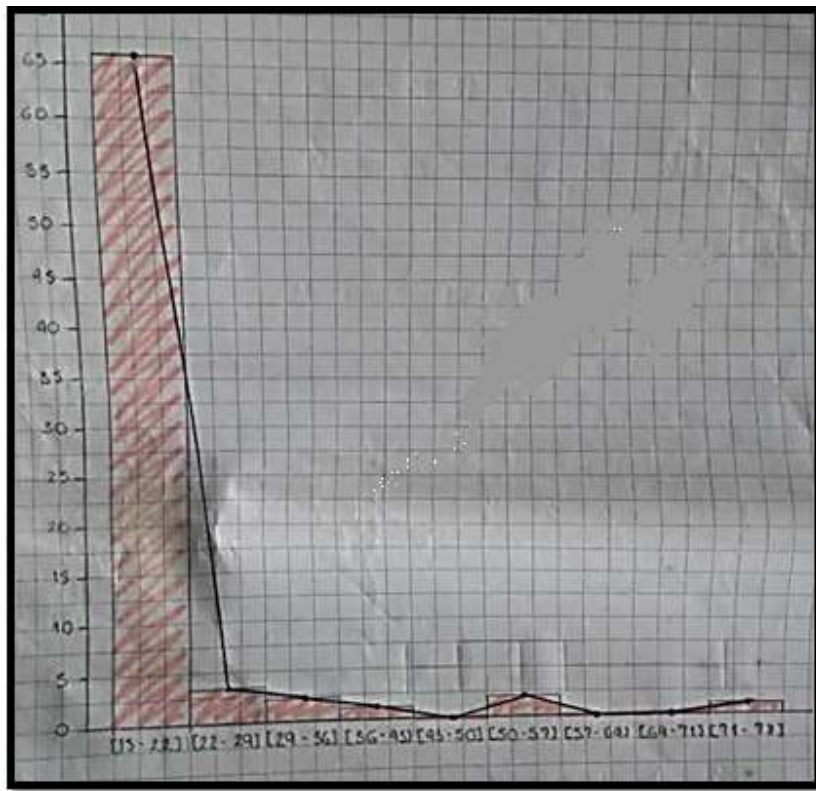


Figura 4-10. Ejemplo 1 Histograma y polígono de frecuencias ítem 6.

## 5. Conclusiones y recomendaciones

Las conclusiones y recomendaciones, que se incluyen a continuación son consecuencia del diseño, implementación y evaluación del proyecto titulado *Estadísticas de las pruebas saber año 2017* con énfasis en el aprendizaje de las medidas descriptivas. A continuación, se presentan las conclusiones para cada uno de los objetivos y las recomendaciones pertinentes para aquellos educadores que deseen poner en práctica esta estrategia didáctica.

### 5.1. Conclusiones sobre los objetivos

Objetivo general. *Desarrollar e implementar una estrategia didáctica para trabajar con los estudiantes de grado décimo las medidas descriptivas mediante la metodología del trabajo por proyectos.*

Este objetivo se cumple con el diseño de la estrategia didáctica titulada *Estadísticas de la prueba Saber año 2017* que se describe en el capítulo 3 y en donde se menciona que este se desarrolla bajo la metodología del trabajo por proyectos apuntando al aprendizaje de las medidas descriptivas, dicho trabajo fue realizado con 120 estudiantes de grado décimo utilizando una sub base de datos de 75 registros, correspondiente a los resultados de la prueba saber del año 2017 del segundo semestre del año. Además, también fue necesario fundamentar el proyecto a partir del marco disciplinar, el marco didáctico y el marco metodológico que se describe en el capítulo 2 del presente trabajo.

Primer objetivo específico. *Identificar los conocimientos previos de los estudiantes con respecto a los conceptos de media aritmética, modas(s), mediana, varianza, desviación estándar, cuartiles y percentiles (Medidas descriptivas).*

Se recolectó datos acerca de los conocimientos previos de los estudiantes respecto a los conceptos de media aritmética, modas(s), mediana, varianza, desviación estándar, cuartiles y percentiles (Medidas descriptivas) mediante la construcción de una prueba diagnóstica con nueve preguntas de selección múltiple y dos preguntas abiertas.

Se identificaron las siguientes dificultades: No interpretar en los gráficos la información pertinente y necesaria para calcular la media aritmética (punto 2 y punto 8), interpretar el valor de la frecuencia absoluta como el equivalente al de la frecuencia relativa (Punto 6 y 7), se reconoce que la mayoría de estudiantes no identifican ciertas representaciones gráficas y no dan cuenta de su interpretación como lo sucedió con el boxplot (punto 9). Así como también sucede que los estudiantes no saben cómo realizar cálculos de la varianza, desviación estándar y las medidas de posición (Punto 10a y 10b)

Con lo anterior se puede concluir que no hay un completo dominio de los conceptos de medidas descriptivas y esto se dificulta por falta de conocimiento de conceptos preliminares.

Segundo objetivo específico. *Proponer ejercicios que conduzcan a la utilización de hojas de cálculo Excel en el análisis de datos, estadística descriptiva.*

Durante el desarrollo del capítulo 3 se hace mención a la sesión 3.1.2 que lleva el título de *descripciones generales de las sesiones previas al desarrollo del proyecto*, en este se menciona la ejecución de 4 sesiones de las cuales dos de ellas, la sesión 3 y 4 dan cuenta de la utilización de hojas de cálculo Excel en donde a los estudiantes se les facilitó y se les dio la opción de trabajar con un computador portátil del colegio. Para la ejecución de estas secciones se utilizaron dos bases de datos obtenidas del DANE, una correspondía al Censo Habitantes de la calle en Bogotá D.C. en el año 2017, y la otra base correspondía a las Estadísticas Vitales de Colombia año 2017. De ambas bases se seleccionaron las sub-bases de 76 y 106 datos respectivamente, y junto con los estudiantes se iban comparando los resultados obtenidos en las hojas de cálculo, de acuerdo a la base de datos correspondiente. Una vez se finalizaron las cuatro sesiones se procedió a realizar la implementación del proyecto en donde a los estudiantes también se le dio la opción de trabajar con hojas de cálculo.

Tercer objetivo específico. *Seleccionar la información, los contextos y las situaciones idóneas para diseñar las actividades que den estructura a la estrategia didáctica mediante el trabajo por proyectos.*

Se diseña un proyecto, utilizando datos correspondientes a los resultados de la prueba Saber 2017 segundo semestre, debido al gran interés mostrado por los estudiantes en los

resultados de la prueba saber. De la base de datos seleccionada, se toma un total de 15 variables en donde por lo menos haya una variable que haga parte de la clasificación (nominal, ordinal, discreta y continua). Los datos seleccionados se consideran pertinentes porque se pueden abordar temas de las medidas descriptivas. Adicionalmente, para contestar algunas preguntas del proyecto era necesario relacionar las 3 variables asignadas para cada uno de los grupos que se describen en la sesión 3.2.3 de manera que el estudiante necesite tratar con información estadística para seleccionar, interpretar y evaluar críticamente la información.

Último objetivo específico. *Implementar y evaluar la estrategia didáctica.*

En el capítulo 4 se hace el análisis de los resultados de la estrategia didáctica, en este se realiza el análisis cualitativo y estadístico de la prueba diagnóstica inicial y de la prueba diagnóstica final después de haberla aplicado, con el propósito de establecer los referentes para evaluar la pertinencia de la estrategia didáctica implementada. A continuación, realizamos el listado de los referentes.

- Referentes del análisis estadístico después de aplicar la estrategia didáctica permiten observar que 9 de los 11 ítems de la prueba diagnóstica inicial se vieron favorecidas con la implementación del proyecto en los cuatro grupos (numeral 4.1.2). Por otra parte, se evalúa la situación de la pregunta 1 que fue la única que desmejoró en los cuatro grupos, aunque las cifras no son alarmantes ya que uno o dos estudiantes desmejoraron por cada curso, si hay que tener en cuenta que fue la única pregunta en la que los resultados desmejoraron en los 4 cursos. Con respecto a la pregunta 3 de la prueba inicial y final se realizaron todos los análisis respectivos, pero no son cifras que muestren total confiabilidad en los resultados, debido a que esta pregunta no se puede comparar ya que no son la misma en la prueba inicial y final. En síntesis, se estima un efecto positivo en la aplicación del proyecto *Estadísticas de las pruebas Saber año 2017*.
- En el análisis cualitativo después de aplicar el proyecto, las actividades de enseñanza aprendizaje muestran una diferencia en los resultados obtenidos, esto se refleja en los porcentajes obtenidos por cada pregunta para cada uno de los cursos, donde al comparar los resultados con respecto a la prueba diagnóstica

luego de aplicar la estrategia didáctica se encuentra que: de las preguntas 4 en adelante se presentan mejores resultados, las preguntas 1 y 2 en algunos cursos presentan desmejora y la pregunta 3 mejoró en algunos grupos pero esta no se tuvo en cuenta en el análisis de resultados.

Es importante recalcar que la implementación del proyecto *Estadísticas de la prueba Saber año 2017* permite realizar un análisis e interpretación de datos estadísticos que no son ajenos al contexto de los estudiantes, esto favorece el entendimiento de los conceptos de medidas descriptivas que son el objeto de estudio en esta propuesta y además hace que los estudiantes se involucren de mejor manera, debido a el interés mostrado hacia los resultados de las pruebas Saber. Además, se evidenció que la mayoría de los estudiantes se vieron muy motivados con la aplicación de la estrategia didáctica y por la reestructuración de la práctica docente en el aula a la que conllevó el desarrollo de este proyecto.

## 5.2. Recomendaciones

Con la ejecución del proyecto *Estadísticas de las pruebas Saber año 2017* se listan las siguientes recomendaciones con el fin de mejorar el desarrollo y la implementación de la estrategia. Además, que estas sirvan como consideraciones importantes para aquellos docentes que deseen poner en práctica esta propuesta de trabajo.

- Una vez analizado los resultados para cada una de las preguntas de la prueba diagnóstica inicial y final se recomienda realizar el ajuste de la pregunta 2 tanto en la prueba inicial como en la prueba final, ya que el cambio realizado en la pregunta hace que los resultados no sean comparables, debido a que está muy alejada de la pregunta que se realizó inicialmente. Para este caso se sugiere escoger la pregunta de la prueba final y ajustar su etiqueta en el eje  $x$  especificando que es costo en euros.
- Cuando se realizó la implementación del proyecto a los estudiantes se les brindó la opción de realizar sus procedimientos mediante la hoja de cálculo de Excel o

realizarlos de manera manual. Al realizar los cálculos de las medidas posición notaron que los resultados obtenidos diferían un poco de los realizados de manera manual esto generó una inquietud, ya que asimilaron que alguno de los dos estaba mal. Debido a esta situación la recomendación es realizar los cálculos de las medidas de posición con interpolación que es la manera en la que lo hace Excel para no generar confusiones en los estudiantes.

- Durante las cuatro sesiones de aprendizaje el tiempo fue bastante limitado y esto dificultó el desarrollo de más ejemplos con las hojas de cálculo de Excel. Para optimizar el tiempo de las sesiones es necesario implementar sesiones que estén netamente dedicadas al uso del programa Excel en el caso de que sea necesario; esto optimizaría la ejecución de las sesiones previas al proyecto como también las de este.

Estas consideraciones están sujetas a diferentes situaciones que se dan dentro de los ambientes de aprendizaje, como lo son los ritmos de aprendizaje.

- Con los datos de los resultados obtenidos se puede desarrollar trabajos futuros que comprenda temas de análisis multivariado como las correlaciones asociadas a las variables.

## A. Anexo: Prueba diagnóstica



### COLEGIO DE LA SALLE: Prueba diagnostica

#### 1. ESTRUCTURA DE LA GUÍA

##### 1.1. Introducción:

Estimado estudiante:

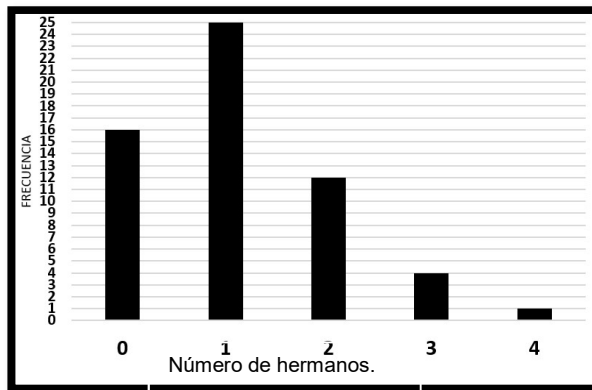
Al desarrollar esta evaluación, aplicarás los conceptos aprendidos en el primer periodo. Se evaluará por ejes y competencias.

Lee con mucha atención cada uno de los siguientes ejercicios, y marca con una X, la letra correspondiente a la respuesta correcta, ten en cuenta que no todas las preguntas son de selección múltiple.

##### 1.2. Desarrollo y aplicación.

Resuelva las preguntas 1,2 y 3 de acuerdo a la siguiente información.

1. El siguiente gráfico muestra la cantidad de hermanos que tienen los estudiantes de grado séptimo.



¿Cuántos estudiantes hay en grado séptimo?

- a. 42
- b. 25
- c. 58
- d. 9

2. ¿Cuál es el promedio de hermanos que tienen los estudiantes de grado séptimo?

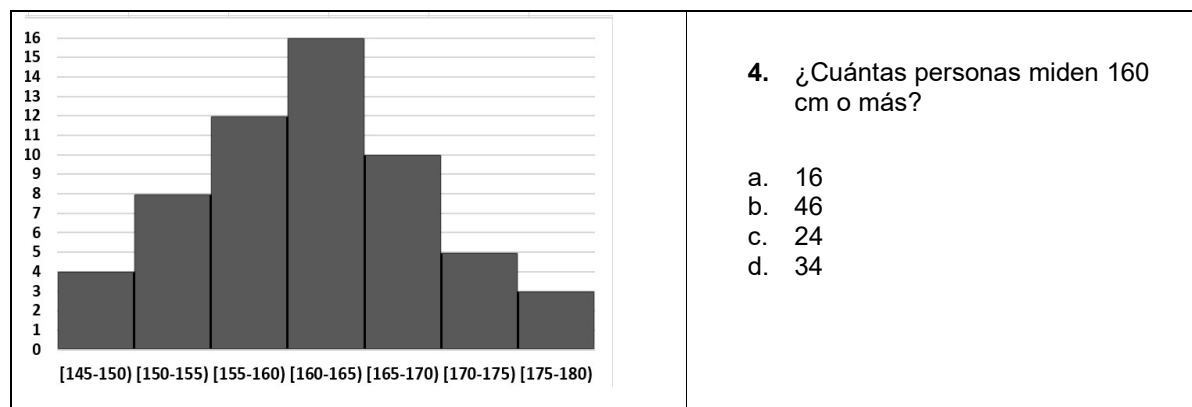
- a. 1,12
- b. 65
- c. 0,84
- d. 1

3. ¿Cuál es la moda?

- a. 2
- b. 3
- c. 4
- d. 1

**Responda las preguntas 4, 5 y 6. De acuerdo a la siguiente información.**

El siguiente histograma muestra las alturas de un grupo de personas.



**4.** ¿Cuántas personas miden 160 cm o más?

- a. 16
- b. 46
- c. 24
- d. 34

**5.** ¿Cuántas personas miden menos de 160 cm?

- a. 12
- b. 24
- c. 46
- d. 16

**Responda las preguntas 6 y 7. De acuerdo al siguiente enunciado.**

La siguiente tabla corresponde a la distribución de frecuencia relativa ( $h_i$ ) de 400 empleados de una empresa, según su edad en años cumplidos.

EDAD	$h_i$
[19-21)	0,05
[21-23)	0,12
[23-25)	0,17
[25-27)	0,28
[27-29)	0,24
[29-31]	0,14

**6.** ¿Cuántos empleados tienen edad mayor o igual que 21 años y menor de 25 años?

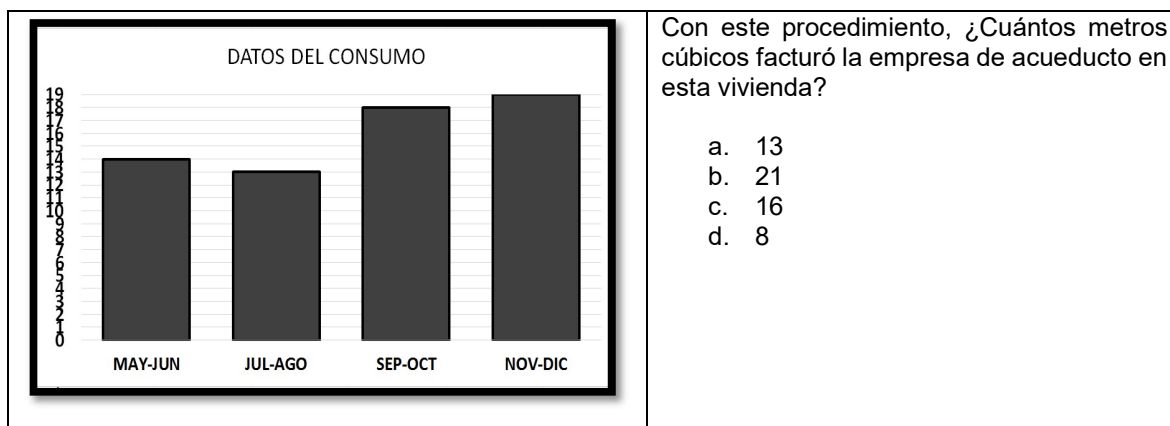
- a. 180
- b. 15.5
- c. 160
- d. 116

**7.** ¿Qué porcentaje de los empleados tienen 25 años o más?

- a. 34%
- b. 66%
- c. 73 %
- d. 28%

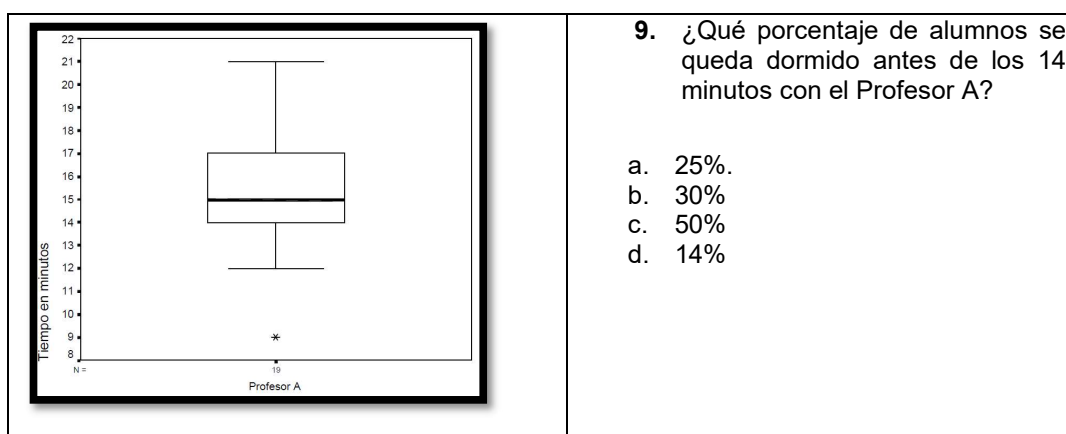
**8.** Para facturar el consumo de agua de una vivienda en la que no se pudo realizar la lectura del contador, la empresa de acueducto promedió el consumo de agua en metros cúbicos de los últimos 4 bimestres. En la siguiente gráfica aparece el consumo en metros cúbicos, de esta vivienda durante cuatro bimestres.





**Responda las preguntas 9 y 10 de acuerdo a la siguiente información.**

Dos profesores (A y B) están interesados en estudiar los hábitos de sueño de los estudiantes en sus clases. Ambos profesores registran el tiempo (en minutos) que demoran en quedarse dormidos sus alumnos desde que empieza la clase. El gráfico muestra los tiempos que demoran en quedarse dormidos los alumnos del profesor A.<sup>4</sup>



**10.** Los datos del Profesor B son los siguientes:

10,5 11,3 11,9 12 12,3 12,3 12,5 12,7 13,4 13,7  
13,8 14,2 14,8 15,1 15,3 16,7 16,8 18,8 20,8

- Calcule la media aritmética (promedio) y la mediana.
- Calcule la varianza, el cuartil 1 y cuartil 3

<sup>4</sup> Tomado de: [www.mat.uda.cl/hsalinas/cursos/2012/eyp1/guia1.pdf](http://www.mat.uda.cl/hsalinas/cursos/2012/eyp1/guia1.pdf)

## B. Anexo: Hojas de códigos. Censo Habitantes de calle Bogotá año 2017, Estadísticas vitales de Colombia año 2017 y Estadísticas de la prueba Saber 11 año 2017

### CODEBOOK ESTADISTICAS DE LA PRUEBA SABER 11 AÑO 2011

ID	ETIQUETA	TIPO DE VARIABLE	PREGUNTA
V1	Estudiante	(indicativo de registro del estudiante)	Directorio
V2	Genero	Cualitativa categórica nominal	GENERO: F: Femenino; M: masculino
V3	Edad	Cuantitativa continua	EDAD
V4	Departamento	Cualitativa categórica nominal	Departamento dónde presenta la prueba
V5	Municipio	Cualitativa categórica nominal	Municipio dónde presenta la prueba
V6	Colegio	Cualitativa categórica nominal	¿Naturaleza del colegio? Oficial No oficial
V7	Jornada	Cualitativa categórica nominal	¿A qué jornada escolar pertenece? Mañana; Tarde; Noche; Sabatina; Completa
V8	Puntuación lc	Cuantitativa continua escala intervalo	Puntaje en lectura crítica.
V9	Percentil lc	Cuantitativa continua escala de razón	Percentil en lectura crítica.
V10	Puntuación mat	Cuantitativa continua escala intervalo	Puntaje en matemáticas.
V11	Percentil mat	Cuantitativa continua escala de razón	Percentil en matemáticas.
V12	Puntuación en cn	Cuantitativa continua escala intervalo	Puntaje en ciencias naturales.
V13	Percentil cn	Cuantitativa continua escala de razón	Percentil en ciencias naturales.
V14	Puntuación ing	Cuantitativa continua escala intervalo	Puntaje en inglés.
V15	Percentil ing	Cuantitativa continua escala de razón	Percentil en inglés.

### CODEBOOK CENSO HABITANTES DE CALLE BOGOTÁ (HCB) AÑO 2017

ID	ETIQUETA	TIPO DE VARIABLE	PREGUNTA
V1	Habitante	(indicativo de registro del HCB)	Directorio
V2	Localidad	Cualitativa categórica nominal	Localidad/Comuna. 1 Usaquén 2 Chapinero 3 Santafé 4 San Cristobal 5 Usme 6 Tunjuelito 7 Bosa 8 Kennedy 9 Fontibón 10 Engativá 11 Suba 12 Barrios Unidos 13 Teusaquillo 14 Los Mártires 15 Antonio Nariño 16 Puente Aranda 17 La Candelaria 18 Rafael Uribe Uribe 19 Ciudad Bolívar
V3	Lugar de la entrevista	Cualitativa categórica nominal	Lugar de la entrevista 1. En la calle (puente, andén, parque, alcantarilla, etc.) 2. En una institución
V4	Años cumplidos	Cuantitativa continua	¿Cuántos años cumplidos tiene usted?
V5	Genero	Cualitativa categórica nominal	GENERO: 1.Hombre; 2. Mujer
V6	Principal razón hc	Cualitativa categórica nominal	¿Cuál fue la PRINCIPAL razón por la que empezó a vivir en la calle? 1. Consumo de sustancias psicoactivas; 2. Por gusto personal; 3. Amenaza o riesgo para su vida o integridad física; 4. Influencia de otras personas; 5. Dificultades económicas; 6. Falta de trabajo; 7. Conflictos o dificultades familiares 8. Abuso sexual; 9. Siempre ha vivido en la calle 10. Víctima del conflicto armado o desplazado 11. Otra"
V7	Sabe leer	Cualitativa categórica nominal	¿Usted sabe leer y escribir? 1. Sí; 2. No
V8	Nivel educativo	Cualitativa categórica ordinal	¿Cuál es el nivel educativo MÁS ALTO alcanzado por usted y el ÚLTIMO AÑO O GRADO aprobado en ese nivel? 1. Preescolar; 2. Básica primaria; 3. Básica secundaria; 4. Media académica, media técnica o normalista; 5. Técnica profesional o tecnológica; 6. Universitario o posgrado; 13. Ninguno
V9	Consigue dinero	Cualitativa categórica nominal	"PRINCIPALMENTE, ¿cómo consigue usted dinero? 1. Limpiando vidrios, cuidando carros, tocando llantas, vendiendo en la calle u otras similares; 2. Cantando, haciendo malabares, cuentería artesanía u otras similares; 3. Carpintería, electricidad, construcción u otras similares; 4. Pidiendo, retacando, mendigando; 5. Recogiendo material reciclable; 6. Como campanero, taquillero, vendiendo o transportando sustancias psicoactivas; 7. Robando o atracando; 8. Ejerciendo la prostitución; 9. Otra"

### CODEBOOK ESTADÍSTICAS VITALES DE COLOMBIA AÑO 2017

ID	ETIQUETA	TIPO DE VARIABLE	PREGUNTA
V1	Neonato	(indicativo de registro del neonato)	Directorio
V2	Cod_dpto	Cualitativa categórica de tipo nominal	"Departamento de Nacimiento Código, según la División Político- Administrativa del DANE.; 05= Antioquia;08= Atlántico;11= Bogotá 13= Bolívar; 15= Boyacá; 17=Caldas; 18= Caquetá; 19=Cauca; 20= Cesar; 23= Córdoba 25=Cundinamarca; 27= Choco; 41= Huila; 44= La guajira; 47= Magdalena; 50= Meta; 52= Nariño; 54= Norte de Santander; 63= Quindío; 66= Risaralda; 68= Santander 70= Sucre; 73= Tolima; 76= Valle del Cauca 81= Arauca; 85= Casanare; 86= Putumayo 88= Archipiélago de San Andrés, Providencia y Santa Catalina; 91= Amazonas; 94= Guainía; 95= Guaviare; 97= Vaupés; 99= Vichada"
V3	Sexo	Cualitativa categórica de tipo nominal	"Sexo del nacido vivo; 1 = Masculino; 2 = Femenino; 3= Indeterminado"
V4	Peso_nac	Cuantitativa continua	"Peso del nacido vivo, al nacer 1 = Menos de 1.000; 2 = 1.000 - 1.499; 3 = 1.500 - 1.999;4 = 2.000 – 2499; 5 = 2.500 - 2.999; 6 = 3.000 – 3499; 7 = 3.500 - 3.999; 8 = 4.000 y más; 9 = Sin información"
V5	Talla_nac	Cuantitativa continua	" Talla del nacido vivo, al nacer 1 = Menos de 20; 2 = 20-29; 3 = 30-39; 4 = 40-49; 5 = 50-59; 6 = 60 y Más; 9 = Sin información"
V6	Mes de nacimiento	Cualitativa categórica tipo nominal	Mes de nacimiento. 1=Enero; 2=Febrero; 3=Marzo;4=Abril; 5=Mayo; 6=Junio; 7=Julio; 8=Agosto; 9=Septiembre; 10=Octubre; 11=Noviembre;12=Diciembre
V7	Atendido por	Cualitativa categórica tipo nominal	"El parto fue atendido por 1 = Médico;2 = Enfermero(a); 3 = Auxiliar de enfermería; 4 = Promotor(a) de salud; 5 = Partera; 6 = Otro persona; 9 = Sin información"
V8	Tipo_parto	Cualitativa categórica tipo nominal	"Tipo de parto de este nacimiento 1 = Espontáneo; 2 = Cesárea; 3 = Instrumentado; 4 = Ignorado; 9 = Sin información"
V9	Edad_madre	Cuantitativa continua	"Edad de la madre a la fecha del parto; 1 = De 10-14 Años; 2 = De 15-19 Años; 3 = De 20-24 Años; 4 = De 25-29 Años; 5 = De 30-34 Años; 6 = De 35-39 Años; 7 = De 40-44 Años; 8 = De 45-49 Años 9 = De 50-54 Años; 99 = Sin información"
V10	Niv_edum	cualitativa categórica de tipo ordinal	"Ultimo nivel de estudio que aprobó la madre Nivel educativo de la madre;1 = Preescolar; 2 = Básica primaria;3 = Básica secundaria; 4 = Media académica o clásica; 5 = Media técnica; 6 = Normalista; 7 = Técnica profesional; 8 = Tecnológica; 9 = Profesional;10 = Especialización; 11 = Maestría; 12 = Doctorado; 13 = Ninguno;99 = Sin información"

## C. Anexo: Proyecto. Estadísticas de las pruebas saber año 2017

Este trabajo tiene como finalidad realizar un estudio acerca de los resultados obtenidos en la prueba Saber del año 2017, para el desarrollo del trabajo se optó por tomar una sub-base de datos con 75 registros tomados de manera aleatoria. A continuación, se describirá como es el trabajo a realizar.

De acuerdo a las variables asignadas realizar la siguiente guía de trabajo. Recuerde que dicho trabajo se debe realizar con los grupos establecidos.

El siguiente recuadro corresponde a las variables asignadas para cada uno de los grupos.

<i>GRUPO A</i>	<i>GRUPO B</i>	<i>GRUPO C</i>	<i>GRUPO D</i>
<i>Variables</i>	<i>Variables</i>	<i>Variables</i>	<i>Variables</i>
<b>V3,V8,V10</b>	V3,V10,V12	V3,V12,V14	V3,V14,V8

Para el desarrollo de esta actividad tenga en cuenta: la base de datos, las tablas de frecuencias y el codebook. La base de datos y las tablas de frecuencia se encuentra en el archivo (BASE DE DATOS ICFES 2017) y el codebook en el archivo; (CODEBOOK ICFES) los archivos se encuentran adjuntos en el correo enviado por plataforma institucional.

A continuación, la actividad a realizar.

1. Determinar para cada una de las variables asignadas la naturaleza (Continua, discreta) y su escala. (razón, intervalo)
2. Escojan dos de las variables y determinen el mínimo, máximo,  $Q_1$ ,  $Q_2$ ,  $Q_3$ ,  $\bar{x}$ ,  $s^2$  y  $s$ .
3. Interpreten  $Q_1$ ,  $\bar{x}$  en cada una de las variables.
4. Discutan sobre la moda en cada una de las variables y coloquen sus conclusiones.
5. Realizar el boxplot con alguna de las variables (No escoger V3) ¿Hay datos atípicos?
6. Realizar el histograma y el polígono de frecuencias de la variable V3.

## D. Anexo: Prueba final.

### 1. ESTRUCTURA DE LA GUÍA

#### 1.1. Introducción:

Estimado estudiante:

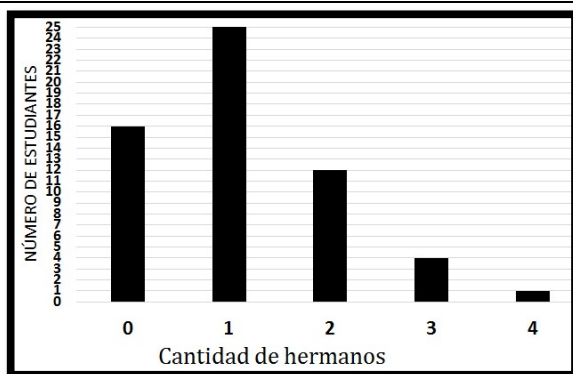
Al desarrollar esta evaluación, aplicarás los conceptos aprendidos. Se evaluará por ejes y competencias.

Lee con mucha atención cada uno de los siguientes ejercicios, y marca con una X, la letra correspondiente a la respuesta correcta, ten en cuenta que no todas las preguntas son de selección múltiple.

#### 1.2. Desarrollo y aplicación.

Resuelva las preguntas 1y 2 de acuerdo a la siguiente información.

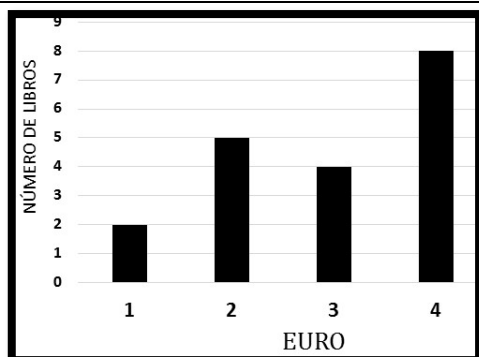
1. El siguiente gráfico muestra la cantidad de hermanos que tienen los estudiantes de grado séptimo.



¿Cuántos estudiantes hay en grado séptimo?

- a. 42
- b. 25
- c. 58
- d. 9

El siguiente gráfico presenta la relación entre costo (euros) de un libro y el número de ejemplares comprados.



2. ¿Cuál es el costo promedio de un libro?

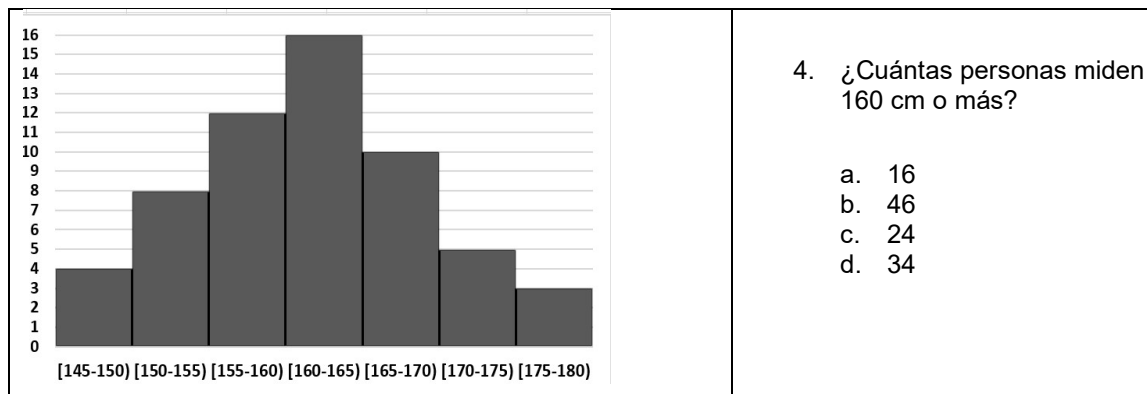
- a. 2.95 Euros.
- b. 56 Euros.
- c. 14 Euros.
- d. 0.53 Euros

3. ¿Cuál es la moda?

- a. 2 euros
- b. 3 euros
- c. 4 euros
- d. 1 euro

**Responda las preguntas 4 y 5. De acuerdo a la siguiente información.**

El siguiente histograma muestra las alturas de un grupo de personas.



4. ¿Cuántas personas miden 160 cm o más?

- a. 16
- b. 46
- c. 24
- d. 34

5. ¿Cuántas personas miden menos de 160 cm?

- a. 12
- b. 24
- c. 46
- d. 16

**Responda las preguntas 6 y 7. De acuerdo al siguiente enunciado.**

La siguiente tabla corresponde a la distribución de frecuencia relativa ( $h_i$ ) de 400 empleados de una empresa, según su edad en años cumplidos.

EDAD	$h_i$
[19-21)	0,05
[21-23)	0,12
[23-25)	0,17
[25-27)	0,28
[27-29)	0,24
[29-31]	0,14

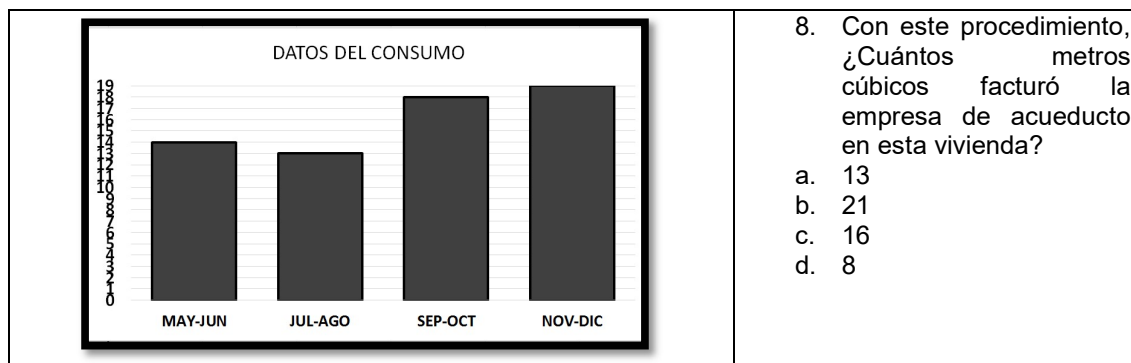
6. ¿Cuántos empleados tienen edad mayor o igual que 21 años y menor de 25 años?

- a. 180
- b. 15.5
- c. 160
- d. 116

7. ¿Qué porcentaje de los empleados tienen 25 años o más?

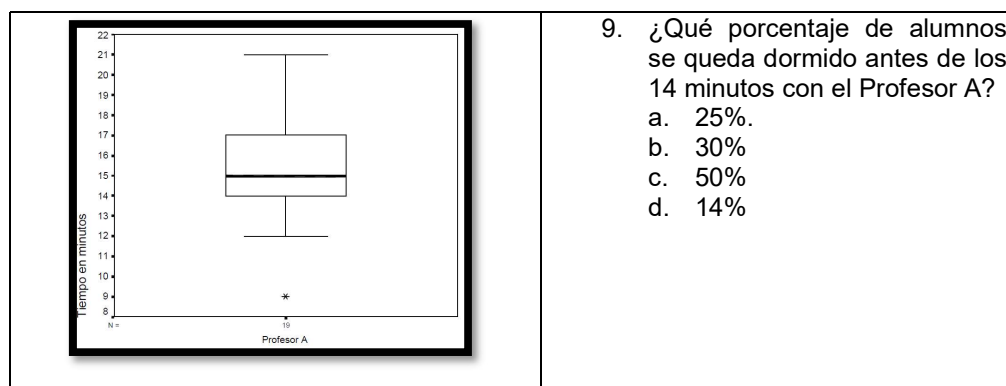
- a. 34%
- b. 66%
- c. 73 %
- d. 28%

Para facturar el consumo de agua de una vivienda en la que no se pudo realizar la lectura del contador, la empresa de acueducto promedió el consumo de agua en metros cúbicos de los últimos 4 bimestres. En la siguiente gráfica aparece el consumo en metros cúbicos, de esta vivienda durante cuatro bimestres.



**Responda las preguntas 9 y 10 de acuerdo a la siguiente información.**

Dos profesores (A y B) están interesados en estudiar los hábitos de sueño de los estudiantes en sus clases. Ambos profesores registran el tiempo (en minutos) que demoran en quedarse dormidos sus alumnos desde que empieza la clase. El gráfico muestra los tiempos que demoran en quedarse dormidos los alumnos del profesor A.<sup>5</sup>



10. Los datos del Profesor B son los siguientes:

10.5; 11.3; 11.9; 12; 12.3; 12.3; 12.5; 12.7; 13.4; 13.7; 13.8; 14.2

- Calcule la media aritmética (promedio) y la mediana.
- Calcule la varianza, el cuartil 1 y cuartil 3



## 6. BIBLIOGRAFIA

- Batanero , C., & Godino, J. C. (2002). *Estocástica y su Didáctica para Maestros*. Granada, España: Universidad de Granada.
- Batanero, C. (2000). Significado y comprensión de las medidas de posición central. *Revista UNO*, 25, 41-58.
- Batanero, C. (2001). *Didáctica de la Estadística* . Granada, España : Departamento de Didáctica de la Matemática, Universidad de Granada .
- Batanero, C., & Díaz , C. (2011). *Estadística con proyectos*. Granada: Facultad de Ciencias de la Educación Universidad de Granada.
- Batanero, C., Díaz, C., & Gea, M. M. (2011). Las estadísticas de pobreza y desigualdad. En C. Batanero, & C. Díaz, *ESTADÍSTICA CON PROYECTOS* (págs. 97-124). Granada: Universidad de Granada.
- Batanero, C., Díaz, C., Contreras, J., & Arteaga, P. (2011). Enseñanza de la Estadística a través de Proyectos. En C. Batanero, & C. Díaz, *Estadística con proyectos* (págs. 1-46). Granada: Universidad de Granada.
- Batanero , C., Díaz, C., Contreras, J. M., & Roa, R. (2013). El sentido estadístico y su desarrollo. *NÚMEROS*, Pag. 7-18.
- Bausela, E. (2004). LA DOCENCIA A TRAVÉS DE LA INVESTIGACIÓN ACCIÓN. *Revista Iberoamericana de Educación*, 1-9.
- Chaparro, H. (2018). *Unidad didáctica para la enseñanza de estadísticos resumen a estudiantes de grado noveno (Tesis de Maestría)*. Bogotá D.C.: Universidad Nacional de Colombia.
- Colmenares, A. M., & Piñero, M. L. (2008). LA INVESTIGACIÓN ACCIÓN Una herramienta heurística para la comprensión y transformación de realidades prácticas socio-educativas . *Laurus*, 14, 96-114.
- Departamento de Didáctica de la Matemática, Facultad de Ciencias de la Educación. (2004). *Didáctica de las Matemáticas para Maestros*. Granada, España: Universidad Granada.
- Estrella, A. (2011). *Historia de la Estadística*. República Dominicana: Revista Actualidad Estadística.

- Estrella, S. (2008). Medidas de tendencia central en la enseñanza básica en Chile. *Revista chilena de educación matemática*, 4, 2-21.
- García-Carmona, A. (2009). La investigación-acción en la enseñanza de la Física: un escenario idóneo para la formación y desarrollo profesional del profesorado. *Latin-American journal of physics Educations*, 388-394.
- Hidalgo, A. (2016). Un Recuento Histórico de la Estadística. *REVISTA SIGMA*, 44-52.
- Hurtado, J. H. (2013). *Breve historia de la estadística (Tesis de maestría)*. Medellín, Colombia : Universidad Nacional de Colombia.
- ICFES. (2016). *Saber 9° Lineamientos para las aplicaciones muestral y censal 2016*. Bogotá D.C.: MEN.
- La Salle Distrito Bogotá . (28 de 09 de 2019). *Colegio de La Salle* . Obtenido de <https://colsalle.edu.co/index.php>
- MEN. (1998). *Lineamiento Curriculares en Matemáticas* . Bogotá : Imprenta Nacional de Colombia.
- MEN. (2006). *Estandares Básicos de Competencias en Matemáticas*. Bogotá: Imprenta Nacional de Colombia.
- MEN. (12 de Abril de 2016). *Derechos Básicos de Aprendizaje v2*. Bogotá D.C.: MEN. Obtenido de [http://aprende.colombiaaprende.edu.co/sites/default/files/naspublic/DBA\\_Matem%C3%A1ticas.pdf](http://aprende.colombiaaprende.edu.co/sites/default/files/naspublic/DBA_Matem%C3%A1ticas.pdf)
- Monroy Saldívar, S. (2008). *Estadística Descriptiva*. México D.F., México: Instituto Politécnico Nacional.
- Rojas, H. (2017). *Construcción, análisis e interpretación de gráficos estadísticos con estudiantes de ciclo IV mediante el trabajo por proyectos*. Bogotá D.C.: Universidad Nacional de Colombia .
- Ruiz, J. D. (2017). *Síntesis de la investigación sobre variabilidad y dispersión en estadística (Trabajo fina de maestría)*. Granada, España: Universidad de Granada.
- Salinas, H. S. (2012). *Guía N°1: Estadística*. Copiapó, Chile: Universidad de Atacama .
- Shepard, L. (2006). LA EVALUACIÓN EN EL AULA. En L. Shepard, & R. Brennan, *Educational Measurement* (págs. 623-646). Colorado: Greenwood Publishing Grpup Inc.

- Sosa Martínez, J. C., Ospina Forero, L. E., & Berdugo Camacho, E. P. (2013). *Estadística Descriptiva y Probabilidades*. Bogotá D.C.: Universidad Externado de Colombia.
- Suárez , M. (2002). Algunas reflexiones sobre la investigación-acción colaboradora en la educación. *Revista Electronica de Enseñanza de las Ciencias*, 1(1), 40-56.